Tissue-specific genetic and epigenetic alterations in *Mlh1* heterozygous tissues

Kul Shanker Shrestha

Genome Stability Group

Research Program in Systems Oncology (ONCOSYS)

Research Programs Unit

Faculty of Biological and Environmental Sciences

Doctoral Programme in Integrative Life Science (ILS)

University of Helsinki

Helsinki, Finland

ACADEMIC DISSERTATION

Doctoral thesis, to be presented for public examination with the permission of the Faculty of Biological and Environmental Sciences of the University of Helsinki, in Porthania PIII, Yliopistonkatu 3, Helsinki on the 8th of October, 2021 at 13:00.

Helsinki 2021

Supervisor:
Liisa Kauppi, PhD
Associate Professor, Department of Biochemistry and Developmental Biology
Research Program in Systems Oncology
Faculty of Medicine, University of Helsinki, Finland

Thesis committee:
Saara Ollila, PhD
Academy Research Fellow, Translational Cancer Medicine Program
Faculty of Medicine, University of Helsinki, Finland

Riikka Kivelä, PhD
Docent, Academy Research Fellow, Stem Cells and Metabolism Research Program
Faculty of Medicine, University of Helsinki, Finland

Reviewers:
Yuri Dubrova, PhD
Professor, Department of Genetics
University of Leicester, United Kingdom

Keijo Viiri, PhD
Adjunct professor, Academy Research Fellow
Faculty of Medicine and Health Technology, Tampere University, Finland

Opponent:
Francesca Cole, PhD
Associate Professor, Department of Epigenetics and Molecular Carcinogenesis
MD Anderson Cancer Center, University of Texas, USA

Custos:
Juha Partanen, PhD
Professor, Molecular and Integrative Biosciences Research Program
University of Helsinki, Finland

ISBN 978-951-51-7519-9 (paperback)

ISBN 978-951-51-7520-5 (PDF)

ISSN 2342-3161 (print)

ISSN 2342-317X (online)

Cover art: Atish Chitrakar © 2021 (atishchitrakar@gmail.com)

Cover layout: Anita Tienhaara

To my family.

"Two roads diverged in a wood, and I—
I took the one less traveled by,
And that has made all the difference."
- Robert Frost

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following articles:

I. **Shrestha, K. S.**, Aska, E. M., Tuominen, M. M., & Kauppi, L. (2021). Tissue-specific reduction in MLH1 expression induces microsatellite instability in intestine of *Mlh1*$^{+/-}$ mice. DNA Repair (Amst), 106, 103178. doi:10.1016/j.dnarep.2021.103178

II. **Shrestha, K. S.**, Tuominen, M. M., & Kauppi, L. (2021). *Mlh1* heterozygosity and promoter methylation associates with microsatellite instability in mouse sperm. *Mutagenesis*. doi:10.1093/mutage/geab010

The articles are referred in the text by their Roman numerals (I and II)

Author's contributions:

Study I: K.S.S designed the experiments, maintained the mouse colony, did tissue collection and processed the samples, extracted DNA, RNA and protein, performed microsatellite instability (MSI) assay, loss of heterozygosity (LOH) assay and methylation specific PCR (MSP) assay, performed quantitative PCR (qPCR) and western blot experiments. K.S.S did the data analysis and interpreted the data. K.S.S. wrote the manuscript.

Study II: K.S.S designed the experiments, maintained the mouse colony, did tissue collection and processed the samples, extracted DNA, performed MSI assay and MSP assay. K.S.S performed the data analysis and interpreted the data. K.S.S. wrote the manuscript.

ABBREVIATIONS

A_{al}	A aligned spermatogonia
A_{int}	A intermediate spermatogonia
A_{pr}	A paired spermatogonia
A_s	A single spermatogonia
BAX	B-cell lymphoma 2 Associated X
cDNA	complementary deoxyribonucleic acid
CG	cytosine-guanine
CHK-1	Checkpoint kinase 1
CMMRD	constitutional mismatch repair-deficiency
CpG	cytosine-phosphate-guanine
CRC	colorectal cancer
DNA	deoxyribonucleic acid
Epas1	endothelial PAS domain-containing protein 1
GI	gastrointestinal
HNPCC	hereditary nonpolyposis colorectal cancer
HRP	horseradish peroxidase
IGFIIR	insulin like growth factor 2 receptor
IgG	immunoglobulin G
IHC	immunohistochemistry
indel	insertion deletion
ISC	intestinal stem cell
LGR5	leu-rich repeat-containing G proteincoupled receptor 5
LOH	loss of heterozygosity
LS	Lynch syndrome
MLH	MutL Homolog
MMR	mismatch repair
mRNA	messenger RNA
MSH	MutS homolog
MSI	microsatellite instability
MSI-H	microsatellite instability-high

MSI-L	microsatellite instability-low
MSP	methylation-specific PCR
MSS	microsatellite stable
n.s.	non-significant
PBL	peripheral-blood leukocytes
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PMS	postmeiotic segregation
PTEN	Phosphatase and tensin homolog
Ptpn20	protein tyrosine phosphatase non-receptor type 20
qPCR	quantitative polymerase chain reaction
RAD-50	RAD50 Recombinase
RIPA	radioimmunoprecipitation assay
RFLP	restriction fragment length polymorphism
SD	standard deviation
SM-PCR	single-molecule PCR
SNP	single nucleotide polymorphism
SSC	spermatogonial stem cell
TA	transit amplifying
TGFβRII	transforming growth factor beta type II receptor
TSS	transcription start site

ABSTRACT

Mismatch repair (MMR) proteins are involved in maintaining genome integrity in somatic and germline cells. Defects in MMR leads to various abnormalities, such as tumorigenesis in Lynch syndrome (LS) and infertility. MLH1 (a central MMR protein) along with other MMR proteins are involved in repair of replication-induced errors arising due to inaccurate DNA polymerase. Compared to other genomic locations, replication-induced errors occur at relatively higher frequency at microsatellites. If unrepaired (due to defective MMR), these errors accumulate and give rise to a cellular phenotype known as microsatellite instability (MSI). MSI is a hallmark of tumors associated with LS and results from complete loss of MMR; LS patients have high incidence of colorectal cancer.

It is still poorly understood why certain tissues in LS, in particular the gastrointestinal (GI) tract, are more vulnerable to MSI-associated tumors. Also, how heterozygosity of MMR genes impact the germline genomic integrity is not known. We hypothesized that certain tissues, in particular those with high proliferating rates, are vulnerable to MMR associated abnormalities compromising microsatellite stability (prior to complete loss of MMR function) in a tissue-specific manner before tumorigenesis. Our aim was to understand how MMR protein levels contribute to MSI *in vivo*.

In the research presented here, we used *Mlh1* heterozygous mice (*Mlh1*$^{+/-}$ mice) as model of LS to investigate how decreased MLH1 levels contribute to tissue-specific MSI, and whether MSI is detectable prior to loss of MMR function and to neoplastic growth. We tested MSI and measured MLH1 levels in primary cells derived from different organs, focusing on the comparison between highly proliferating small intestine and low proliferating spleen. Further, we tested MSI in sperm cells of *Mlh1*$^{+/-}$ mice. In addition, we studied the association of loss of heterozygosity (LOH) and *Mlh1* promoter methylation to tissue-specific MLH1 expression and MSI. The studies were conducted at 4- and 12-month time point.

We discovered that highly proliferating normal tissues (small intestine and sperm) of *Mlh1*$^{+/-}$ mice display MSI which increases with age, while low proliferating spleen was microsatellite stable. Further, *Mlh1*$^{+/-}$ small intestine showed sporadic decrease in

MLH1 levels which associated with the observed MSI, while *Mlh1*$^{+/-}$ spleens showed expected MLH1 expression (i.e. approximately 50% of wildtype expression). We observed soma-wide *Mlh1* promoter methylation in a subset of *Mlh1*$^{+/-}$ mice; these mice were the most vulnerable to MLH1 expression level decrease and to MSI in the small intestine, while MLH1 expression in other somatic tissues remained unaffected. In brief, we showed that normal small intestine of *Mlh1*$^{+/-}$ mice is particularly susceptible to MLH1 depletion giving rise to MSI long before neoplasia. Further, we demonstrated that *Mlh1*$^{+/-}$ sperm exhibit MSI which associates with germline *Mlh1* promoter methylation.

1. REVIEW OF THE LITERATURE

1.1. DNA Mismatch repair (MMR) in eukaryotes

MMR proteins, a class of DNA repair proteins, are responsible for the surveillance and maintenance of DNA integrity. MMR proteins are involved in the repair of replication-induced DNA damage such as base-base mismatches, small insertion-deletions (indels) which arise due to the infidelity of DNA polymerases, and in processing of induced (in mitotic cells) and cell-programmed (in meiotic cell) DNA double-stranded breaks which are the most lethal type of DNA damage (Baker et al., 1996; McCulloch & Kunkel, 2008; Pray, 2008; Shao et al., 2004; Thomas et al., 1991). In eukaryotes, DNA polymerases has an error rate 1×10^{-7} per nucleotide (or less) on average *in vivo* (Ganai & Johansson, 2016), most but not all of these DNA damages are repaired by various cellular mechanisms, these mechanisms include proofreading function of DNA polymerases (during replication), DNA mismatch repair system (post-replication repair) and homologous recombination (double strand break repair) (Pray, 2008). Despite the repair, during every cell division an average of 5×10^{-11} and 1.8×10^{-10} mutations per nucleotide is introduced into human and *Mus musculus* genome, respectively (Drake, Charlesworth, Charlesworth, & Crow, 1998), which reflects incomplete repair functions.

The MMR system in eukaryotes involves several mismatch repair proteins: MutS homologs (MSH2, MSH3, MSH4, MSH5, and MSH6) and MutL homologs (MLH1, MLH3, PMS1, and PMS2) (Wei, Kucherlapati, & Edelmann, 2002). Heterodimers MSH2-MSH6 (MutSα) and MSH2-MSH3 (MutSβ) recognize mismatches between the nascent (i.e. newly synthesized) and the parental strand (MutSα recognizes base-base mismatch and 1-bp indels, and MutSβ recognizes 1-4 bp indels) (Boland & Goel, 2010; Wei et al., 2002) (**Figure 1**). Heterodimers (primarily) MLH1-PMS2 (MutLα) and MLH1-MLH3 (MutLγ) are involved in repair of the mismatches in mitotic cells (Chen et al., 2005; Wei et al., 2002) (**Figure 1**). In meiotic cells, MutLγ is involved in repair of double strand breaks that form during meiotic recombination (Rogacheva et al., 2014).

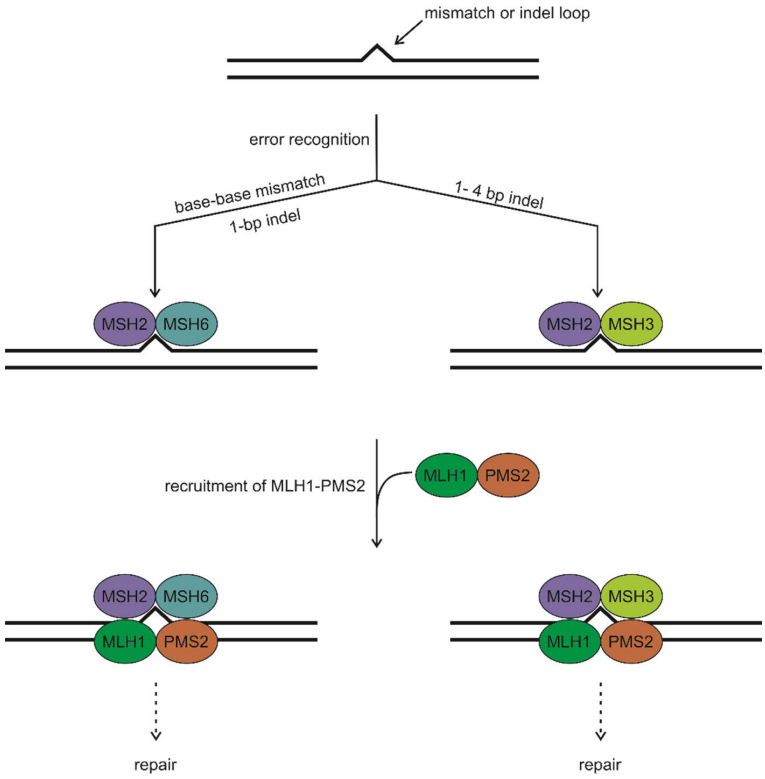


Figure 1. Involvement of MMR proteins in repair of replication errors. Mismatches and indel loops are recognized by MSH2-MSH6 (MutSα) and MSH2-MSH3 (MutSβ). MutSα and MutSβ recruit MutLα (heterodimer MLH1-PMS2) which activates the downstream repair events.

1.2. Microsatellites and microsatellite instability (MSI)

Human and mouse genomes comprise of 50%-70% and up to 45% repetitive elements, respectively (de Koning, Gu, Castoe, Batzer, & Pollock, 2011; Lu et al., 2020). One class of repetitive elements is tandem repeats. As the name implies, the genomic sequence in these repetitive elements is arranged in tandem repeat arrays. The tandem repeats makeup 3% and 3-5% of the human and mouse genome, respectively (Komissarov, Gavrilova, Demin, Ishov, & Podgornaya, 2011; Treangen & Salzberg, 2011). A subclass of tandem repeats is satellite DNA; microsatellites are the most abundant type of satellite DNA (Dumbovic, Forcales, & Perucho, 2017) (**Figure 2**). Microsatellites are defined as tandem repeat DNA sequence up to a total of 100 nucleotides in length, consisting of 1-6 bp (or 1-10 bp) repeat units (there is not yet a consensus on the upper threshold of the repeat-unit size of a microsatellite) (Bois et al., 1998; Fan & Chu, 2007; Richard, Kerrest, & Dujon, 2008; Yauk, Dubrova, Grant, & Jeffreys, 2002).

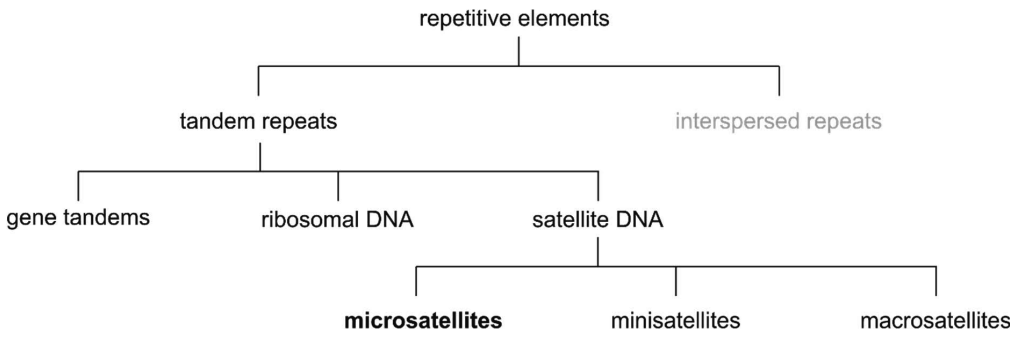


Figure 2. Classification of tandem repeat DNA loci. Repetitive elements in DNA can be divided into tandem and interspersed repeats. Shown is the sub-classification of tandem repeats. Microsatellites belong to the class of satellite DNA.

Due to their repetitive nature, microsatellites are highly prone to replication errors ("replication slippage") by DNA polymerase (error rate: 1×10^{-6} to 1×10^{-2} per locus per generation) (Eckert & Hile, 2009; Ellegren, 2000, 2004; Fan & Chu, 2007). As a consequence, DNA loops are formed at the microsatellites. In the presence of proficient MMR, these DNA loops are repaired post-replication. In the absence of MMR, in the subsequent round of DNA replication, these DNA loops give rise to *de novo* insertion/deletion of repeat units at the microsatellites. This cellular phenotype is referred to as microsatellite instability (MSI). Insertion at microsatellite occurs if the DNA loop is formed on the newly synthesized strand, and deletion occurs if the DNA loop is formed in the template strand (Fan & Chu, 2007) (**Figure 3**).

Various factors determine the mutability of a microsatellite, and some microsatellites can be 100-fold more unstable than others (Eckert & Hile, 2009). The primary determining factor of mutability of a microsatellite is its intrinsic features such as repeat unit size, length of repeat array and sequence composition (Eckert & Hile, 2009; Fan & Chu, 2007; Shah, Hile, & Eckert, 2010). Also, the genomic location of microsatellite effects its mutability: exonic microsatellites accumulate less mutations as they are well protected by MMR compared to intronic or intergenic microsatellites (Frigola et al., 2017). Microsatellites are relevant study targets for two reasons: first, they are sensitive indicators of MMR defects (Mead et al., 2007) and genotoxicity (Beal et al., 2015; Fennelly, Wright, & Plumb, 1997), and second, when intragenic and mutated, they can be the root cause of neoplasms and other diseases (Duval & Hamelin, 2002; Markowitz et al., 1995).

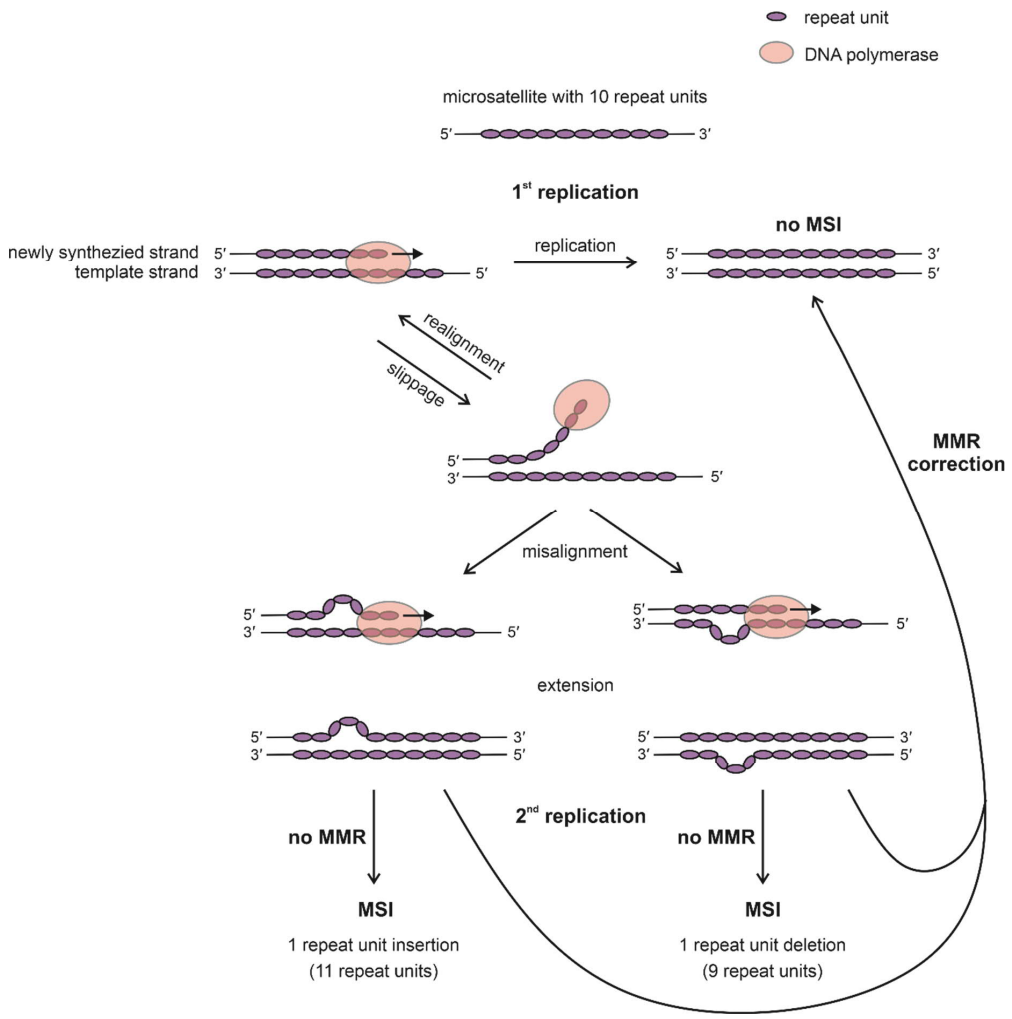


Figure 3. Schematic diagram of microsatellite instability (MSI). DNA polymerase are prone to slippage at microsatellites. Misalignment of strands after slippage creates a DNA loop. In the absence of MMR, the DNA loop remains unrepaired, and in the subsequent DNA replication, the DNA loop is either copied (if the loop is formed on the nascent strand) or omitted (if the loop is formed on the template strand), giving rise to expansion or contraction of the microsatellite, respectively. The outcome is referred to as MSI. Figure adopted and modified from (Fan & Chu, 2007).

1.3. MMR, MSI and cancer

1.3.1. Lynch syndrome (LS)

Lynch syndrome (LS) is an autosomal-dominant cancer syndrome that accounts for approximately 2%-5% of all colorectal cancer (CRC) (Duraturo, Liccardo, De Rosa, & Izzo, 2019; Hampel et al., 2005; Wijnen et al., 1998). LS individuals have more than 80% and 43%-60% lifetime risk of developing CRC and endometrial cancer, respectively (Burt, 2000). These individuals display early onset of CRC (average age of 44 years) (Burt, 2000). Other types of cancers in LS tumor spectrum accounting to a lifetime cancer risk of 1-13% includes stomach/gastric cancer, ovarian cancer, hepatobiliary tract cancer, urinary tract cancer, small bowel cancer, brain/central nervous system cancer and sebaceous neoplasm (Watson & Riley, 2005).

The LS individuals have a mono-allelic germline mutation in one of the MMR genes, predominantly in *MLH1* or *MSH2* (in approximately 90% of LS cases), and occasionally in other MMR genes (*MSH2, MSH6, PMS2* or *PMS1*), *MSH6* and *PMS2* mutation accounts for 10% of LS cases (Bonadona et al., 2011; Lynch & de la Chapelle, 2003). Earlier, LS was known as hereditary nonpolyposis colorectal cancer (HNPCC) because cancer incidence in these patients was thought to be confined to nonpolyposis CRC. Over course of time, there were observations of incidence of other cancers in the HNPCC patients, thus, the condition of monoallelic germline mutation in MMR genes is currently referred to as LS. However the term HNPCC is not obsolete.

International conventions have set specific guidelines for LS diagnosis. These guidelines have been updated/revised periodically to make LS diagnosis more accurate, including better stratification of LS CRC cases from sporadic CRC cases. An international convention held in Amsterdam in 1990 issued the first guideline for HNPCC diagnosis named as Amsterdam criteria I. This guideline was limited to diagnosis of nonpolyposis CRC. In 1990's, HNPCC was better understood (example: discovery of novel cellular phenotypes associated to HNPCC, and observation of other CRC and non-GI tumors in HNPCC patients). In 1998, the Amsterdam criteria I was revised to Amsterdam criteria II. For better LS diagnosis, later the Bethesda guidelines were established, and revised in 2003. Amsterdam criteria II, and revised Bethesda

guidelines are listed in **Table 1.** A notable addition to the Bethesda guidelines was inclusion of MSI criterion.

Table 1. Guidelines for LS diagnosis. Amsterdam criteria II (Vasen, Watson, Mecklin, & Lynch, 1999) is listed above, and revised Bethesda guideline is listed below (Umar, Risinger, Hawk, & Barrett, 2004)

Amsterdam criteria II (1998) *
There should be at least three relatives with HNPCC-associated cancer, with one relative being a first-degree relative of the other two
At least two successive generations should be affected
At least one individual should be diagnosed before the age of 50 years
Familial adenomatous polyposis CRC should be excluded
Tumors should be verified by pathological examination

Revised Bethesda guideline (2003) **
CRC should be diagnosed before the age of 50 years
Presence synchronous or metachronous colorectal or other HNPCC-related tumor regardless of age
CRC with a MSI-high (MSI-H) morphology diagnosed before the age of 60 years
CRC with one or more first-degree relatives with CRC or other HNPCC-related tumors. One of the cancers must have been diagnosed before the age of 50 years (with exception to adenoma, which must have been diagnosed before the age of 40 years)
CRC with two or more relatives with CRC or other HNPCC-related tumor, regardless of age.

* all criteria need to be fulfilled, ** only one criteria needs to be fulfilled

1.3.2. LS (suspected) patients

There are individuals, categorized as LS (suspected) patients, who have similar clinicopathological phenotypes as LS patients (i.e. early onset of colorectal and endometrium cancer, and MMR deficient and MSI-H tumors), however they do not have germline deleterious MMR mutations. Instead, the LS (suspected) patients have a germline epigenetic MMR defect (primarily *Mlh1* promoter methylation) (Niessen et al., 2009). The LS (suspected) cases sum up to 0.6-13% of the total LS patients

(Niessen et al., 2009). The LS (suspected) patients show soma-wide *Mlh1* methylation phenotype, i.e. *Mlh1* promoter methylation in multiple tissues. These tissues include: blood, peripheral blood leukocytes (PBLs), buccal mucosa, hair follicles, saliva, colon mucosa, endometrium, gastric mucosa, small bowel, skin fibroblasts, and sperm (Damaso et al., 2019; Gazzoli, Loda, Garber, Syngal, & Kolodner, 2002; Goel et al., 2011; Hitchins, 2013; Hitchins et al., 2007; Kanaya et al., 2003; Miyakura et al., 2004; Morak et al., 2008; Niessen et al., 2009; Suter, Martin, & Ward, 2004).

1.3.3. MSI phenotype in LS: a hallmark of LS-associated cancer

Almost all (more than 90%) of LS- associated tumors show MSI (Aaltonen et al., 1993; Aaltonen et al., 1994; Pedroni et al., 1999; Thibodeau, Bren, & Schaid, 1993). MSI phenotype was first reported in LS tumors in 1993 (Aaltonen et al., 1993), Aaltonen et. al. showed the RER+ (replication error) phenotype by assaying MSI at seven microsatellite loci. They proposed RER+ phenotype as a valuable tool for LS-associated cancer diagnosis, especially incase of incomplete family CRC history, and for better stratification of the LS-associated CRC versus the sporadic CRC, which now-a-days is implemented in the Bethesda guidelines as the MSI assay. As per Bethesda guidelines, a standard MSI panel includes 5 microsatellites: BAT25, BAT26, D2S123, D5S346 and D17S250; depending on MSI at two or more, one or none of the 5 microsatellites, a tumor is classified as MSI-high (MSI-H), MSI-low (MSI-L) or microsatellite stable (MSS), respectively (Umar, Boland, et al., 2004).

Tumors in LS are MMR-deficient and display high levels of MSI (Hemminki et al., 1994), which led to the notion that MSI in LS-associated CRC follows Knudson's two-hit hypothesis of tumorigenesis (Knudson, 1971). According to this model, the "first hit" is inherited as a germline defect in an MMR gene, and the "second hit" makes the remaining functional MMR allele defective (either by a genetic (LOH and/or deleterious mutations) or by an epigenetic (gene silencing by promoter methylation) process) thereby completely losing the MMR function, and as consequence cancer cells exhibit the MSI mutator phenotype (Hemminki et al., 1994; Peltomaki, 2014).

MSI, as reported by the microsatellite markers, confer the global genome instability. However, for a microsatellite to instigate tumorigenesis MSI should occur at

microsatellite located inside a gene; genes that contain intragenic microsatellites are called MSI-target genes. MSI in MSI-target genes may lead to frameshift mutations, giving rise to dysfunctional proteins that affect various regulatory pathways such as cell cycle, cell proliferation, apoptosis and DNA repair, giving selective advantage to cells and making them tumorigenic (Duval & Hamelin, 2002; Vilkki et al., 2002). The first MSI-target genes to be described in MSI-H CRC, and the ones extensively studied are: tumor suppressor genes *TGFβRII* and *IGFIIR*, apoptosis regulator *BAX*, and MMR genes *MSH6* and *MSH3* (Duval & Hamelin, 2002; Markowitz et al., 1995). In recent years, more MSI-target genes with possible roles in tumorigenesis has been described, including gene encoding tumor suppressor protein PTEN, MMR protein MLH3, DNA repair protein RAD50 and DNA damage response protein CHK1 (Duval & Hamelin, 2002). Because of the severity of intragenic mutations (which could lead to catastrophic consequences, for example tumorigenesis), intragenic regions (primarily exons) are more efficiently protected by MMR compared to other genomic regions (Frigola et al., 2017). This preferential MMR activity is due to enriched histone marks H3K36me3 (Histone H3 trimethylation at lysine 36) in the exons which facilitate recruitment of MutSα (MMR heterodimer responsible for mismatch-recognition) (Frigola et al., 2017).

1.4. DNA methylation and Loss of heterozygosity (LOH)

DNA methylation is a major epigenetic modification which preferentially occurs at a CG dinucleotide (a cytosine nucleotide followed by a guanine nucleotide in a 5′ to 3′ orientation), termed as CpG sites (i.e. C-phosphate-G), or simply CG sites. CpG sites are said to have undergone methylation when a methyl group is covalently attached to the C5 position of cytosine nucleotide to form 5-methylcytosine (5-mC) (Holliday & Pugh, 1975).

Regions of the genome with high density of CpG (GC content of more than 50%, and an observed/expected ratio of CpG to GpC greater than 0.6) are called CpG islands, CpG islands are 300-3000 bp in size (Antequera & Bird, 1993; Long, Smiraglia, & Campbell, 2017). There are approximately 45,000 and 37,000 CpG islands per haploid genome in human and in mouse, respectively (Antequera & Bird, 1993). 55.9% and

46.9% of all genes in human and mouse, respectively, are associated with CpG islands (Antequera & Bird, 1993). Within a gene, abundance of CpG islands is the highest in the core promoter. Methylation of CpG islands in the promoter region of a gene is known to suppress gene expression, specifically hypermethylation of CpG islands in the core promoter (promoter region nearest to the start codon containing the DNA polymerase binding site, TATA box and transcription start site (TSS)) is known to silence a gene (Saxonov, Berg, & Brutlag, 2006). DNA methylation is essential for development and tissue-specific gene expression, however, aberrant promoter methylation of key regulatory genes, e.g. tumor suppressor genes, is pathogenic and can lead to several diseases including cancer (Jones & Baylin, 2007).

LOH is a common genetic mechanism leading to inactivation of tumor suppressor genes in an inherited cancer predisposition with mono-alleleic germline mutation (eg. MMR genes in LS). The only functional allele of tumor suppressor gene is lost due to LOH making the gene nonfunctional which leads to tumorigenesis (Ryland et al., 2015). LOH are large structural alterations encompassing loss of part of a chromosome to loss of entire chromosome (Thiagalingam et al., 2001). Mechanisms leading to LOH include genetic recombination, break-induced replication, gene conversion, mitotic nondisjunction and chromosomal segment loss due to deletion events (Thiagalingam et al., 2001).

1.5. MMR and fertility

1.5.1. Spermatogenesis

Spermatogenesis is the process of differentiation of spermatogonia (spermatogonial stem cells (SSCs)) into haploid spermatozoa (the male gametes). In mice, undifferentiated A_{single} spermatogonia (A_s) undergo up to 10 rounds of mitosis to produce primary spermatocytes, and the primary spermatocytes undergoes a two-step meiosis cell division (meiosis I and II) to give rise to sperm cells (**Figure 4**). With this division rate, a single A_s should give rise to 4096 sperm cells. Around 50% of cells undergo apoptosis in the spermatogenesis process, but even so, 40 million sperm are generated per gram of testis tissue per day in mice (Fayomi & Orwig, 2018). In

comparison, in human, the undifferentiated SSC undergoes five divisions (three mitotic and two meiotic) producing approximately 4.4 million sperm cells per gram of testis tissue per day (Fayomi & Orwig, 2018).

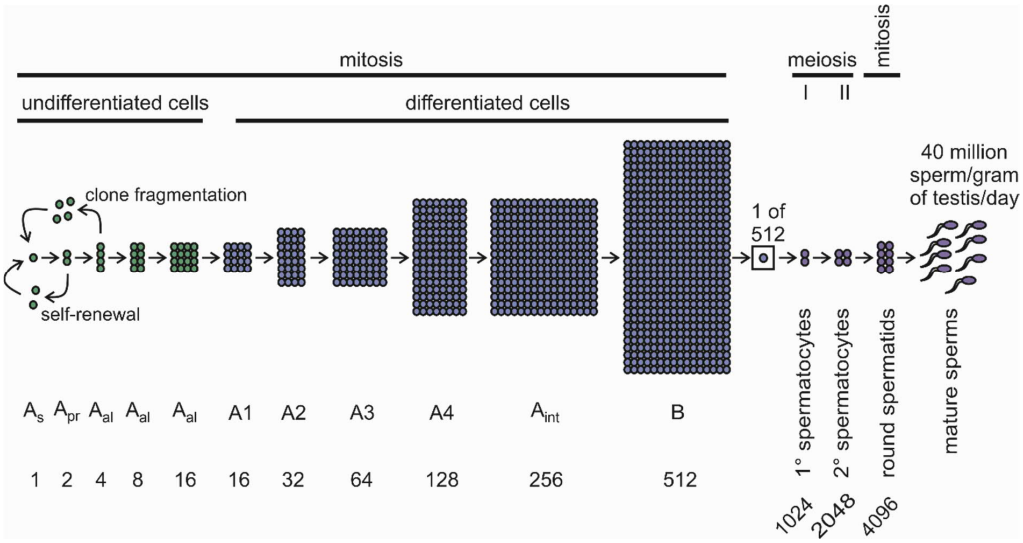


Figure 4. Schematic diagram of spermatogenesis in mice. During spermatogenesis, a single spermatogonial stem cell (A_{single} spermatogonia) undergoes up to 12 transit cell divisions to give rise to roughly 4000 spermatids. Numbers below the spermatogonial sub-stages indicate the cell count at that sub-stage. A_s, A_{single} spermatogonia; A_{pr}, A_{paired} spermatogonia; A_{al}, $A_{aligned}$ spermatogonia; A1, A1 spermatogonia; A2, A2 spermatogonia; A3, A3 spermatogonia; A4, A4 spermatogonia; A_{int}, Intermediate spermatogonia; B, B spermatogonia. Figure modified from (Fayomi & Orwig, 2018).

1.5.2. Role of MMR in spermatogenesis

Spermatogenesis involves a high rate of cell proliferation up to 12 transit amplifying divisions (10 mitotic, followed by two meiotic cell divisions) of a single spermatogonial stem cell. With this high rate of cell division, maintaining the genomic integrity

throughout spermatogenesis is crucial, especially as any DNA aberration in the sperm cells can be inherited to the next generation. Therefore, the role of MMR in spermatogenesis is critical. During spermatogenesis, MMR proteins are involved in post-replicative DNA repair in pre-meiotic germline cells, and in meiotic recombination and gametogenesis in meiotic cells. As in somatic cells, in pre-meiotic germline cells MMR proteins MLH1, PMS2, MSH2, MSH3 and MSH6 are involved in repairing indel mutations and single base pair mismatches. In meiotic cells, MMR proteins MLH1, MLH3, MSH4 and MSH5 are essential to complete meiotic crossovers and for gametogenesis (Baker et al., 1996; de Vries et al., 1999; Edelmann, Cohen, et al., 1999; Gunes, Al-Sadaan, & Agarwal, 2015; Kneitz et al., 2000; Lipkin et al., 2002; Wei et al., 2002). In the germline, MMR defects lead to failure in spermatogenesis and oogenesis and dysfunctional sperm and oocytes, and can even lead to infertility in human and mice (Baker et al., 1996; Gunes et al., 2015; Lipkin et al., 2002; Mukherjee, Ridgeway, & Lamb, 2010).

1.6. Mouse model for MMR

Mouse models have been extensively used to study eukaryotic MMR and MMR-associated phenotypes. Several mouse models resembling the MMR deficiency in human have been created from the late 1990's to early 2000's. These includes mouse models for MutS homologs: *Msh2* (Lin et al., 2004), *Msh3* (de Wind et al., 1999), *Msh5* (de Vries et al., 1999; Edelmann, Cohen, et al., 1999), *Msh6* (de Wind et al., 1999; Yang et al., 2004) and mouse models for MutL homologs: *Mlh1* (Baker et al., 1996; Edelmann et al., 1996), *Mlh3* (Lipkin et al., 2002), *Pms1* (Prolla et al., 1998), *Pms2* (Baker et al., 1995). The table below lists widely used MMR (gene knock-out) mouse models along with severity status of MSI, cancer incidence and the effect on fertility.

Table 2. MMR knock-out mouse models and their severity status of molecular (MSI) and tissue-level (cancer and fertility) phenotypes. Modified from (Lee, Tosti, & Edelmann, 2016)

gene		molecular phenotype	observed phenotype		reference
knock-out		MSI	cancer	fertility	
MutS	*Msh2* $^{-/-}$	severe	severe	no effect	(de Wind, Dekker, Berns, Radman, & te Riele, 1995; Egashira et al., 2002; Reitmair et al., 1995)
	Msh3 $^{-/-}$	moderate	moderate	no effect	(de Wind et al., 1999; Edelmann et al., 2000)
	Msh4 $^{-/-}$	no effect	no effect	severe	(Kneitz et al., 2000)
	Msh5 $^{-/-}$	no effect	no effect	severe	(de Vries et al., 1999; Edelmann, Cohen, et al., 1999)
	Msh6 $^{-/-}$	moderate	moderate	no effect	(Edelmann et al., 1997)
MutL	***Mlh1*** $^{-/-}$ *	severe	severe	severe	(Edelmann, Yang, et al., 1999; Kawate et al., 1998; Prolla et al., 1998)
	Mlh3 $^{-/-}$	moderate	severe	moderate	(Chen et al., 2005)
	Pms1 $^{-/-}$	nominal	no effect	moderate	(Prolla et al., 1998)
	Pms2 $^{-/-}$	severe	severe	severe	(Prolla et al., 1998)

* *Mlh1* mice with different *Mlh1* gene dosage (i.e. *Mlh1*$^{+/+}$, *Mlh1*$^{+/-}$ and *Mlh1*$^{-/-}$) were used in studies presented in this thesis.

1.6.1. *Mlh1*$^{+/-}$ mice as a model for Lynch syndrome

Mlh1 mice are good animal models to study the relationship between MMR gene dosage and the subsequent genomic instability and tumor incidence. *Mlh1* heterozygous (*Mlh1*$^{+/-}$) and *Mlh1* homozygous (*Mlh1*$^{-/-}$) mice carry germline mono- and bi-allelic deleterious mutation in the *Mlh1* gene, respectively, which mimics the germline MMR status of LS syndrome (see section 1.3.1) and constitutional mismatch repair-deficiency (CMMRD) syndrome in human, respectively (Edelmann et al., 1996;

Tabori et al., 2017). CMMRD patients carry germline bi-alleleic deleterious mutation in one of the MMR genes, primarily *PMS2* and *MSH6*.

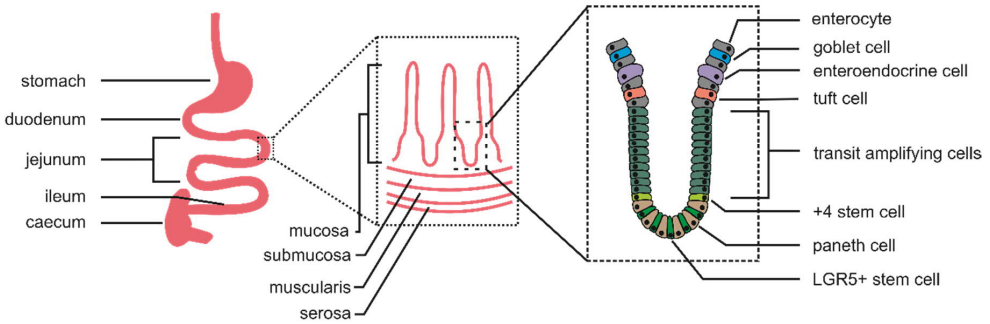
Mlh1$^{-/-}$ mice has been extensively studied for the MSI phenotype and tumor incidence. The tumor spectrum in *Mlh1*$^{-/-}$ mice spans primarily GI tract and hematological malignancies, and skin cancers (Prolla et al., 1998). Median age of onset of cancer in *Mlh1*$^{-/-}$ mice is before 7 months (Edelmann, Yang, et al., 1999; Tabori et al., 2017). In human, the median age of onset of first cancer in CMMRD patient is 7.5 years (versus 67 years in a healthy person). Most common tumors in CMMRD patients are brain tumors, followed by GI-tract and hematologic malignancies, CMMRD patients also develop LS-associated malignancies (Tabori et al., 2017).

LS patients have high incidence and early onset of CRC and endometrial cancer (see section 1.3.1). In mouse model of LS (*Mlh1* $^{+/-}$ mice) the median age of onset of cancer is before 9.8 months, and tumorigenesis affects primarily GI tract and hematological tissues (Edelmann, Yang, et al., 1999). However, the tumor incidence in *Mlh1*$^{+/-}$ mice is considerably low, with a wide range reported in various studies from 1% to 32% (1.1% (1 of 88 *Mlh1*$^{+/-}$ mice) (Tokairin et al., 2006), 9% (3 of 32 *Mlh1*$^{+/-}$ mice) (Fu et al., 2009) and 32% (7of 22 *Mlh1*$^{+/-}$ mice) (Edelmann, Yang, et al., 1999). In GI tract of *Mlh1* $^{+/-}$ mice, tumors were observed in the duodenum, jejunum, and colon (Edelmann, Yang, et al., 1999). In *Mlh1*$^{-/-}$ mice, tumors are observed in all parts of the GI tract (jejunum and ileum being the most frequent sites for tumorigenesis) (Edelmann, Yang, et al., 1999; Prolla et al., 1998). In addition, both male and female *Mlh1*$^{-/-}$ mice are infertile, while *Mlh1*$^{+/-}$ mice are fertile (Baker et al., 1996).

1.6.2. The mouse small intestine

The small intestine in mouse is approximately 30 cm in length stretching from the stomach on the proximal end to caecum on the distal end. It can be divided into three distinct regions: proximal duodenum, middle jejunum and distal ileum (**Figure 5A**). The mucosa of small intestine consists of columnar epithelium which forms villi and intestinal crypts. Villi are finger-like absorptive projections facing towards the lumen, composed of terminally differentiated cells (enterocytes, goblet cells and enteroendocrine cells). Crypts consist of undifferentiated LGR5+ intestinal stem cells (ISCs) which reside at the very base of the crypt and are sandwiched between differentiated Paneth cells. Crypts also contain transit amplifying (TA) cells (Li & Jasper, 2016) (**Figure 5A**). Actively dividing LGR5+ ISCs either self-renew, or generate progenitor TA cells. The highly proliferating TA cells undergo two to five divisions and gradually differentiate and migrate from crypt to villi (Barker, 2014; Li & Jasper, 2016) (**Figure 5B)**. There are reserve stem cells which are localized at the +4 position from the base of the crypt called the +4 stem cells, these cells can also self-renew, proliferate and generate different cell lineages of the small intestine, however they remain quiescent and are activated only in case of intestinal injury (Barker, 2014). Small intestine is a highly proliferating tissue. In mice, more than 300 million new epithelial cells are generated every day to compensate for the high rate of cell death on the villi (with a turnover rate of 3-5 days). This is repeated several hundred times during the lifespan of an inbred mouse (which has average life span of 2-year) (Barker, 2014), and all these cells originate from ISCs. Tumor incidence in the GI tract of MMR-deficient mice preferentially affects the small intestine (Prolla et al., 1998) unlike in the colon in the human. Thus, in MMR mouse models, small intestine is the relevant tissue to study the cellular and molecular phenotypes related to GI tumorigenesis in human.

A.



B.

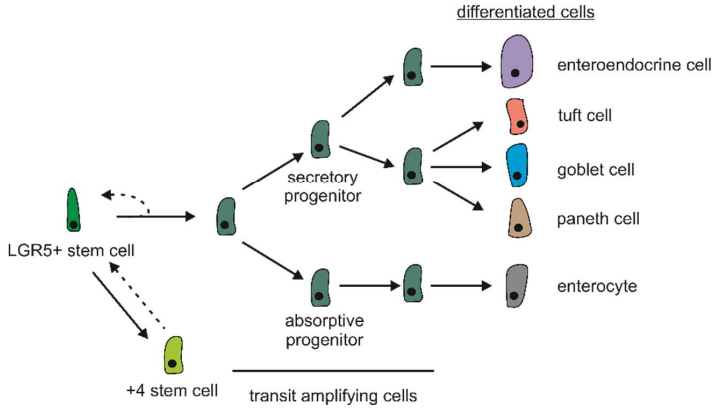


Figure 5. Murine small intestine and the associated intestinal stem cell (ISC) lineages. (A) Anatomy of the mouse small intestine. Mouse small intestine can be divided into three regions: duodenum, jejunum and ileum. A cross-section of small intestine shows multiple layers: namely serosa, muscularis, submucosa and mucosa. Mucosa (the internal lining) consists of villus-crypt units. The crypt contains ISCs and Paneth cells, villus consists of terminally differentiated cells. **(B) ISC lineages** (modified from (Sancho, Cremona, & Behrens, 2015)) ISCs either self-renew or give rise to highly proliferating progenitor transit amplifying (TA) cells. TA cells terminally differentiate into specialized cells.

2. AIM OF THE STUDY

The principal aim of the study was to understand how MMR protein levels contribute to MSI *in vivo*.

Specific aims of the projects were:

Study I. To test tissue-specific *Mlh1* haploinsufficiency in *Mlh1*$^{+/-}$ somatic tissues with various proliferation rates *in vivo*.

Study II. To investigate the effect of *Mlh1* heterozygosity and *Mlh1* promoter methylation on sperm MSI.

3. MATERIALS AND METHODS

Table 3. List of methods used in the original publications

Method	Publication
***Mlh1* genotyping**	I, II
Single-molecule MSI analysis by PCR	I, II
RNA expression analysis using qPCR	I
Western blot analysis	I
Immunohistochemistry (IHC) and IHC image analysis	I
methylation analysis by methylation-specific PCR (MSP)	I, II
Loss of heterozygosity (LOH) analysis by PCR	I
Statistical analysis	I, II

3.1. Mice and genotyping

Mlh1 mice (B6.129- *Mlh1*^{tm1Rak}, strain 01XA2, National Institutes of Health, Mouse Repository, NCI Frederick) (Edelmann et al., 1996) were used in the studies. In B6.129- *Mlh1*^{tm1Rak} mice, exon 2 of the 129S1/SvImJ allele is replaced by a neo cassette making the allele nonfunctional.

Mlh1 genotyping was performed using DNA lysate from earpieces. DNA lysate was prepared using 100 µl of lysis buffer (50 mM KCl, 10 mM Tris-HCl pH 8.3, 2.5 mM MgCl2, 0.1 mg/mL gelatin, 0.45% NP-40, 0.45% Tween20) supplemented with 20 µg proteinase K. Cells were lysed overnight at 56°C, and the next day proteinase K was inactivated by boiling for 10 minutes. 0.5 µl of DNA lysate was used for genotyping PCR, PCR was performed in Platinum green hot start PCR master mix (Invitrogen, Carlsbad, CA) using *Mlh1* genotyping primer published at the Frederick national laboratory's mouse repository website, namely: M001: 5'-TGTCAATAGGCTGCCCTAGG-3', M002 :5'-TGGAAGGATTGGAGCTACGG-3', M003: 5'-TTTTCAGTGCAGCCTATGCTC-3'. Primer combination M001/M002

amplified the knockout allele, and M001/M003 amplified the wildtype allele which had PCR product size of 500 bp and 350 bp, respectively. The following PCR program was used: 2 min at 94°C, 38 cycles of 30 sec at 98°C, 30 sec at 55°C and 45 sec at 72°C, followed by 5 min at 72°C.

3.2. Mouse husbandry, sample collection and preparation

The mice were bred and maintained following Animal Experiment Board in Finland and Laboratory Animal Centre of the University of Helsinki regulations and guidelines. Mice were humanely euthanized using CO_2 inhalation and cervical dislocation. Tissues (except of small intestine) were harvested, snap-frozen and stored at - 80°C until further use. The small intestine was flushed with cold PBS, opened longitudinally, rinsed with cold PBS and inspected for tumors visible to the naked eye. A 3 cm tissue piece was cut from the center of small intestine (approximately 15 cm from the pyloric sphincter), which comprises the jejunum section of the small intestine. This piece of jejunum was inspected for any obvious tumor-like outgrowths using a stereoscope, thereafter snap-frozen and stored at - 80°C. Frozen tissues were used for subsequent DNA, RNA and protein analysis. Only normal (tumor-free) tissues were used for the experiments.

3.3. Single-molecule MSI analysis by PCR

3.3.1. DNA preparation

For somatic tissues, approximately 2-5 mg of tissue was used for DNA extraction. Each tissue piece was finely chopped and homogenized using a 20G needle and syringe, thereafter, DNA was extracted using AllPrep DNA/RNA Mini Kit (Qiagen, Hilden, Germany) according to manufacturer's instruction. For sperm cells, prior to DNA extraction, we performed differential lysis for sperm cell enrichment, to obtain a sperm cell purity of over 95%. This was accomplished using previously published protocol (Zhang, Monckton, Siciliano, Connor, & Meistrich, 2002). Sperm cells were collected from cauda epididymides. The sperm DNA was extracted using DNeasy Blood &

Tissue Kit (Qiagen) following manufacturer's instruction. DNA quantification was performed using a Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA). A working DNA dilution was made by diluting the stock DNA to a dilution of (theoretical) 5 DNA molecule concentration, i.e. 30 pg/µl and 15 pg/µl (considering weight of 3 pg/haploid mouse genome (Laird, 1971)) for diploid and haploid genome, respectively in 5 mM Tris-HCl (pH 7.5) supplemented with 5 ng/µl carrier herring sperm DNA (Thermo Fisher Scientific). For each working DNA dilution to be analyzed, using a dilution series, we determined the concentration that yielded 50% PCR success rate for the microsatellites assayed. By Poisson approximation this PCR success rate equates to approximately one amplifiable molecule per positive PCR reaction (Beal et al., 2015; Yauk et al., 2002).

3.3.2. Microsatellites assayed

We tested MSI at three microsatellite loci: two mononucleotide repeats A27 and A33 (Kabbarah et al., 2003), and one dinucleotide repeat D14Mit15 (Edelmann et al., 1996). The dilution series experiment followed by Poisson approximation (see section 3.3.1) was performed separately for each of the three microsatellites. A27 is an intergenic microsatellite located approximately 2 kb downstream of the *Epas1* gene, A33 is an intronic microsatellite within *Epas1* gene (between exons 2 and 3), and D14Mit15 is an intergenic microsatellite at 40 kb distance from the nearest gene *Ptpn20*.

3.3.3. PCR and capillary electrophoresis

PCR was performed using Q5 High-Fidelity DNA Polymerase system (New England Biolabs, Ipswich, MA) (supplemented with 1 ng/µl carrier herring sperm) in a 10 µl reaction volume. PCR primers used in the MSI assay were as follows: for A27, A27_F 5′ 6-FAM- TCCCTGTATAACCCTGGCTGACT 3′ and A27_R 5′ GCAACCAGTTGTCCTGGCGTGGA 3′, for A33, A33_F 5′ VIC-TACAGAGGATTGTCCTCTTGGAG 3′, A33_R 5′ GCTGCTTCACTTGGACATTGGCT 3′, for D14Mit15, D14Mit15_F 5′ NED TTGGCTGCTCACTTGCAG 3′ and D14Mit15_R 5′ TTACCCTCCCCATAACTCCC 3′ (Edelmann et al., 1996; Kabbarah et

al., 2003). A concentration of 0.5 µM, 0.2 µM and 0.1 µM of forward and reverse primers were used for A27, A33 and D14Mit15, respectively. Two microsatellites, A33 and D14Mit15, were duplexed in a single PCR reaction, and a separate PCR reaction was run for A27. For the A33 and D14Mi15 duplexed PCR, we used the following PCR program: 30 sec at 98°C, 35 cycles of 10 sec at 98°C, 30 sec at 66°C, 5 sec at 72°C, followed by 2 min at 72°C. For A27, the following PCR program was used: 30 sec at 98°C, 35 cycles of 10 sec at 98°C, 30 sec at 70°C, 5 sec at 72°C, followed by 2 min at 72°C. Capillary electrophoresis, using ABI3730xl DNA Analyzer (Thermo Fisher Scientific), was performed to analyze the DNA fragments.

3.3.4. Fragment analysis and mutant scoring

Fragment analysis was performed using Fragman R package (Covarrubias-Pazaran, Diaz-Garcia, Schlautman, Salazar, & Zalapa, 2016). Stringent criteria for true microsatellite signal calling was used, hence the mutation rates reported here are likely a conservative estimate. Mutant scoring criteria can be read in **Study II**.

For each microsatellite assayed, MSI was separately scored for insertions and deletions. We considered the step-wise mutation model for MSI which assumes mutations at microsatellites occur as single repeat-unit changes (Ohta & Kimura, 1973). MSI rate was calculated as follows:
MSI rate = total no. of single repeat-unit shifts observed/total DNA molecules analyzed

In study II, MSI rate is presented as percentage.

3.4. RNA expression analysis using quantitative PCR

RNA used for qPCR was extracted in parallel with genomic DNA, from the same tissue piece, using AllPrep DNA/RNA Mini Kit (Qiagen) as per manufacturer's instructions. Concentration of RNA was measured using NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA). cDNA was synthesized from 500 ng of RNA using SuperScript VILO cDNA Synthesis Kit (Invitrogen). 15 ng of cDNA was used to perform qPCR, using SsoAdvance Universal SYBR Green Supermix system (Bio-Rad, Hercules, CA) using following PCR program: 30 sec at 95°C, 40 cycles of 10 sec at 95°C, 30 sec at 60°C, and 5 sec at 72°C. The PCR plates were read for SYBR green signal after every PCR cycle. Primers used for qPCR were as follows: for *Mlh1*, Mlh1_F 5´ GGGAGGACTCTGATGTGGAA 3´ and Mlh1_R 5´ AGAGCTTGGTCTGGTGCTGT 3´ (amplicon size: 216 bp), and for Beta-actin, beta-actin_F 5´ AGACTTCGAGCAGGAGATGG 3´ and beta-actin_R 5´ AGGTCTTTACGGATGTCAACG 3´ (amplicon size: 310 bp). cDNA from *Mlh1*$^{+/+}$ and *Mlh1*$^{-/-}$ tissues were used as positive and negative control for *Mlh1* expression, respectively. Gene expression data of *Mlh1* was normalized to beta-actin. Data was analyzed using Bio-Rad CFX Maestro (version 1.1) and is presented as % wildtype.

3.5. Protein expression analysis

3.5.1. Western blot analysis

The frozen tissue pieces were thawed on ice, mechanically homogenized and incubated in RIPA buffer (supplemented with protease inhibitor cocktail (Roche, Basel, Switzerland)) for 30 minutes on ice. Thereafter, the protein lysate was centrifuged for 10 minutes at 14000 rpm (at 4°C), the supernatant was collected to a separate tube and stored at - 80°C until further use. Protein concentration was measured using Pierce™ BCA™ protein assay kit (Thermo Fisher Scientific). Western blotting was performed with 30 µg of denatured protein. The protein was run on a 4–20% Mini-Protean TGX gel (Bio-Rad). Using Trans-Blot Transfer Pack (Bio-Rad), protein was

transferred from the gel to 0.2 µm nitrocellulose membrane. Ponceau staining (5 minutes at room temperature) was performed to confirm complete protein transfer to the membrane. 5% milk in PBS supplemented with 1µl/ml Tween-20 (Thermo Fisher Scientific) was used to block the membrane. Blocking was done for 1 hour at room temperature. After blocking, the membranes were incubated overnight at 4°C with primary antibodies against the candidate proteins. The next day, for protein detection, the membranes were incubated with infrared secondary antibodies for 1 hour at room temperature and then scanned with LI-cor Odyssey FC system (LI-COR, Nebraska, USA). Image analysis was performed using LI-cor Image Studio lite (version 5.2). The following antibodies were used for protein detection: primary antibodies against MLH1 (ab92312, Abcam, dilution 1:1000) and Beta-actin (A5441, Sigma, dilution 1:5000), secondary antibodies IRDye 800CW (926-32211, Li-cor, dilution 1:5000) and IRDye 680RD (926-68070, Li-cor, dilution 1:5000). The MLH1 protein signal intensities were normalized to Beta-actin signal intensity.

3.5.2. Immunohistochemistry (IHC) and IHC image analysis

The “Swiss roll” technique was used to embedded the small intestine into paraffin blocks (Moolenbeek & Ruitenberg, 1981). Prior to embedding, the small intestine was flushed with cold 1xPBS, fixed overnight with 4% paraformaldehyde, and cut open longitudinally. 4µm thick sections of the Swiss roll were used for IHC. For IHC, antigen retrieval was performed by heating the slides in 10mM citrate buffer (pH 6) for 20 minutes, followed by staining with antibody against MLH1 (anti- MLH1 antibody, cat. no. ab92312, clone EPR3894, Abcam, Cambridge, UK, dilution 1:1500 dilution). Thereafter, slides were stained with BrightVision Poly HRP goat anti-rabbit IgG (cat. no. DPVR55HRP, ImmunoLogic, Duiven, The Netherlands) and counterstained with hematoxylin-eosin. 3DHistech Panoramic 250 FLASH II digital slide scanner (3DHistech, Budapest, Hungary) was used to scan the slide at 20X magnification. CaseViewer (version 2.2) was used to visualize the slides. Image quantification was done using IHC Profiler in ImageJ (Varghese, Bukhari, Malhotra, & De, 2014). Approximately, 50 villus-crypt units per jejunum from five randomly selected sites was quantified for the staining intensity.

3.6. Methylation-specific PCR (MSP) and LOH assay

MSP (Herman, Graff, Myohanen, Nelkin, & Baylin, 1996) was performed to investigate methylation status of the *Mlh1* promoter. 500 ng of genomic DNA was bisulfite-converted using EZ DNA Methylation-Direct Kit (Zymo Research). 1µl of bisulfite-converted DNA was used for MSP, MSP was performed using 2xZymo Taq premix system (Zymo research). Two parallel PCR reaction, one with primers specific to methylated CpG site and other with primers specific to unmethylated CpG site was performed. For study I following primer were used: primers specific to methylated CpG site: forward primer 5′ GGTGTACGAAGTTATTTTTATTTTAGTC 3′ and reverse primer 5′ ACCCAACGATACCTAATAATAAAACC 3′ (Zymo Research, Irvine, CA, catalog no. D2012), and primers specific to unmethylated CpG site: forward primer 5′ GGTGTATGAAGTTATTTTTATTTTAGTT 3′ and reverse primer 5′ ACCCAACAATACCTAATAATAAAACC 3′. For study II previously published primers were used for MSP (Fraga et al., 2004), which were as follow: primers specific to methylated CpG site: forward primer 5′ GAATTTGAGCGTGAGGAGTTC 3′, reverse primer 5′ TAACCGACCGCTAAATAACTTCC 3′ and primers specific to unmethylated CpG site: forward primer 5′ AGAATTTGAGTGTGAGGAGTTT 3′ and reverse primer 5′ CCAACCACTAAATAACTTCCC 3′. Following PCR program was used in study I: 10 sec at 95°C, 40 cycles of 30 sec at 95°C, 45 sec at 58°C, and 60 sec at 72°C, and final extension for 7 min at 72°C. For study II, following PCR program was used: 10 sec at 95°C, 40 cycles of 30 sec at 95°C, 30 sec at 62°C, and 60 sec at 72°C, and final extension for 7 min at 72°C. PCR products were run on an agarose gel and visualized under UV in presence of ethidium bromide. Methylation status was scored qualitatively based on presence or absence of PCR product in PCR reactions with primers specific to methylated CpG site.

LOH at *Mlh1* was tested using restriction fragment length polymorphism (RFLP) assay targeting the 3′ and the 5′ ends of the *Mlh1* gene, which covered exon 1 and 19 of the *Mlh1* gene, respectively. DNA from $Mlh1^{+/+}$ and $Mlh1^{-/-}$ tissues were used as wildtype and knockout allele controls, respectively. 50 ngs of DNA was used for PCR. PCR was performed using Phusion High-Fidelity DNA Polymerase system (Thermo Fisher

Scientific) and the following PCR condition: 30 seconds at 98°C, 35 cycles of 10 sec at 98°C, 20 sec at 70°C and 120 sec at 72°C, and final extension of 10 min at 72°C. For exon 1 region, following primers were used: forward primer 5´ GGCTTACCTGCCAGCACAACC 3´ and reverse primer 5´ CCGTGTGCATAATGGGAAACC 3´, and for the exon 19 region following primers were used: forward primer 5´ GAGTATGCCAGTAGCTGGGAG 3´ and reverse primer 5´ CAGTTCAAAGATCGGGCAAG 3´. PstI (New England Biolabs) and VspI (Thermo Fisher Scientific) restriction enzymes were used to digest PCR product from exon 1 and exon 19 region, respectively. Digested PCR products were run in 1.5% agarose gel stained with ethidium bromide. The banding pattern of the digested *Mlh1*$^{+/-}$ samples were compared to that of the *Mlh1*$^{+/+}$ and *Mlh1*$^{-/-}$ controls. The sample was considered to have undergone LOH if the banding pattern of *Mlh1*$^{+/-}$ samples were identical to the banding pattern of *Mlh1*$^{-/-}$ control.

4. RESULTS

4.1. Contribution of *Mlh1* gene dosage, age and proliferative capacity of tissues to MSI (Study I-II)

To investigate how the proliferative capacity of the *Mlh1*$^{+/-}$ tissues effect the tissue-specific MSI, we performed highly sensitive SM-PCR based MSI assay in highly proliferating small intestine (specifically, jejunum) and sperm, and in low proliferating spleen. We tested MSI at three microsatellite loci, two mononucleotide repeat: A27 and A33, and a dinucleotide repeat: D14Mit15. The same tissues from age-matched *Mlh1*$^{+/+}$ and *Mlh1*$^{-/-}$ mice were used as MMR-proficient and MMR-deficient controls, respectively (with the exception of sperm cells as *Mlh1*$^{-/-}$ do not produce sperm cells due to failure in spermatogenesis, see section 1.5.2). This experimental set-up gave us an added advantage of comparing MSI across specific tissues in the *Mlh1*-proficient and *Mlh1*-deficient system *in-vivo*. Experiments were carried out in young and old adult mice (4-and 12-months) which gave us the opportunity to assess contribution of age to tissue-specific MSI. In addition, the MSI assay was also performed in intestinal tumor from a 12-month old *Mlh1*$^{-/-}$ mice.

4.1.1. Highly proliferating *Mlh1*$^{+/-}$ tissues show MSI at mononucleotide repeats; dinucleotide repeat was stable (Study I-II)

In *Mlh1*$^{+/-}$ mice, compared to age-matched *Mlh1*$^{+/+}$ mice, we observed increase in deletions at mononucleotide repeats in small intestine and in sperm (2.2-fold ($p = 0.007$) and 2.7-fold ($p= 0.001$) at 4-month, and 4.8-fold ($p=0.001$) and 2.3-fold ($p=0.003$) at 12-month time point, respectively). No such increase was seen in spleen (see **Table 4** for deletion rates). We saw an age-dependent increase in deletions in *Mlh1*$^{+/-}$ small intestine and in *Mlh1*$^{+/-}$ sperm from 4-month to 12-month time point (1.5-fold ($p= 0.006$) and 1.3-fold ($p=0.008$), respectively), while *Mlh1*$^{+/-}$ spleen did not show

such age-dependent increase (see **Table 4** for deletion rates). The deletions observed in *Mlh1*$^{+/-}$ tissues were almost exclusively single repeat unit (i.e.1bp) deletions (**Figure 6**).

Unlike mononucleotide repeats, the dinucleotide repeat was stable in all the three *Mlh1*$^{+/-}$ tissues analyzed at both time points (See **table 5** for MSI rates). While *Mlh1*$^{+/-}$ sperm and *Mlh1*$^{+/-}$ spleen showed almost no deletions at D14Mit15, a small number of deletions were detected at D14Mit15 in *Mlh1*$^{+/-}$ small intestine, and this number increased with age. As at the mononucleotide repeats, the deletions detected at dinucleotide repeat in *Mlh1*$^{+/-}$ tissues were single repeat unit (i.e. 2bp) deletions (**Figure 6**).

One outlier *Mlh1*$^{+/-}$ mouse was identified. This 4-month old *Mlh1*$^{+/-}$ mouse showed substantially higher deletions at mononucleotide repeats in all the three tissues analyzed (32.8%, 7.9%, and 4.9% in small intestine, sperm and spleen, respectively), compared to other age-matched *Mlh1*$^{+/-}$ mice (6-fold, 1.7-fold, and 2-fold higher, respectively). We also observed slightly elevated deletions at dinucleotide repeat (3.9%) in small intestine of this mouse. Unlike deletions observed in other *Mlh1*$^{+/-}$ small intestine (i.e. single repeat unit deletions), in this mouse a high proportion of 2 repeat unit deletions was also observed (both at mono- and dinucleotide repeats) (**Supplementary figure 3/I**), while sperm and spleen showed only single repeat unit deletions. This mouse was excluded from all the statistical analysis (both in **Study I and II**) and is referred to as “outlier” mouse here forward.

4.1.2. Deletions increase with age and proliferative capacity of tissues in *Mlh1*$^{-/-}$ mice (Study I)

Compared to age-matched *Mlh1*$^{+/+}$ tissues (which were either microsatellite stable or showed only baseline level of deletions), *Mlh1*$^{-/-}$ tissues showed substantially higher deletions both at mono- and dinucleotide repeats. For mononucleotide repeats, this

increase was 20.1-fold (p=0.0001) and 14.4-fold (p=0.0001) higher at 4-months, and 58.2-fold (p=0.0001) and 24.1-fold (p=0.0001) higher at 12-months in small intestine and spleen, respectively, and for the dinucleotide repeat, 56-fold (p=0.0001) and 23.5 (p=0.0001) higher at 4 months and 41-fold (p=0.0001) and 34.5-fold (p=0.0001) higher at 12-months in small intestine and spleen, respectively) (see **Table 4** and **5** for MSI rates). We also saw an age-wise increase in deletions at both mono- and dinucleotide repeats in *Mlh1*$^{-/-}$ tissues from 4 months to 12 months. In *Mlh1*$^{-/-}$ small intestine this increase was significant (2-fold (p= 0.01) and 2.2-fold (p= 0.01) at mono- and dinucleotide repeat, respectively), while in *Mlh1*$^{-/-}$ spleen it was not (1.6- fold (p= n.s.) and 1.5-fold (p=n.s.) at mono- and dinucleotide repeat, respectively) (see **Table 4** and **5** for MSI rates).

Unlike in *Mlh1*$^{+/+}$ and *Mlh1*$^{+/-}$ tissues, where we observed almost exclusively single repeat unit deletions, the deletion size shifts in the *Mlh1*$^{-/-}$ tissues both in small intestine and in spleen were larger. At 4 months, size shift involved up to 5- and 3-repeat units at mono-, and 1- and 2-repeat unit at dinucleotide repeats in *Mlh1*$^{-/-}$ small intestine and spleen, respectively; single repeat unit deletions were the most prominent type of mutant allele (**Figure 6**). The size shifts further increased with age in *Mlh1*$^{-/-}$ small intestine, while *Mlh1*$^{-/-}$ spleen showed similar size-shits as at 4-months, at 12-months, *Mlh1*$^{-/-}$ small intestine displayed up to 7- and 3- repeat-unit deletions at mono- and dinucleotide repeats, respectively (**Figure 6**).

Deletions were even more frequent in *Mlh1*$^{-/-}$ intestinal tumor (**Figure 6**), compared to macroscopically normal small intestine of 12-month old *Mlh1*$^{-/-}$ mice, intestinal tumor from a 12-month old *Mlh1*$^{-/-}$ mice showed 1.5-fold, and 1.3-fold more deletions at mono- and dinucleotide repeat, respectively. Similar to the *Mlh1*$^{-/-}$ small intestine at 12 month time point, the *Mlh1*$^{-/-}$ intestinal tumor displayed up to 7- and 3- repeat-unit deletions at mono- and dinucleotide repeats, respectively. Most prominent mutant alleles in the *Mlh1*$^{-/-}$ intestinal tumor were 3 and 4-bp deletions at A27, and 1 and 4-bp at A33, while in *Mlh1*$^{-/-}$ small intestine it was 1 and 2 bp deletions at both A27 and

A33). At D14Mit15, though the prominent allele was 1 repeat-unit deletion in both tumor and in *Mlh1*$^{-/-}$ normal intestine, the frequency of 1 repeat-unit deletion further increased in the tumor compared to the *Mlh1*$^{-/-}$ normal intestine (**Figure 6**).

In general, mononucleotide repeats were more unstable than dinucleotide repeat. Deletions at mononucleotide repeats were approximately 4- to 5-fold higher than at the dinucleotide repeat both in the *Mlh1*$^{+/-}$ and the *Mlh1*$^{-/-}$ tissues (see **Table 4** and **5** for MSI rates).

4.1.3. Opposite dynamics of insertions versus deletions at mononucleotide repeats in proliferating tissues (Study I-II)

While *Mlh1*$^{+/+}$ tissues displayed only baseline levels of deletions (on average 2% and 0.2% at mono- and dinucleotide repeats, respectively), we observed insertional burden at the mononucleotide repeats (particularly at A33) in *Mlh1*$^{+/+}$ tissues. At 4 months, A27 and A33 showed on average 2%, 4%, 3% and 6%, 8%, 10% insertions in *Mlh1*$^{+/+}$ small intestine, sperm and spleen, respectively. In *Mlh1*$^{+/+}$ jejunum insertions further increased with age (for A27 p= n.s., A33 p= 0.01), while in *Mlh1*$^{+/+}$ spleen and sperm insertion frequencies were similar at 4- and 12- month time points (see **Table 4** for MSI rates).

In general, in small intestine and sperm, insertions at mononucleotide repeats (in-particular at A33) tended to decrease with decreasing *Mlh1* gene dosage, especially at the 12 month time-point, while in spleen no such decrease was observed (see **Table 4** for MSI rates). Insertions at mononucleotide repeats further decreased in *Mlh1*$^{-/-}$ intestinal tumor to 1.6%. In *Mlh1*$^{-/-}$ tissues, unlike deletions, insertions were relatively less affected by age. At the dinucleotide repeat, *Mlh1*$^{+/+}$ and *Mlh1*$^{+/-}$ tissues showed negligible insertions at both time points, while *Mlh1*$^{-/-}$ small intestine and *Mlh1*$^{-/-}$ spleen showed 2.8% and 2.3%, and 3.8% and 2.7% insertions at 4- and 12-months, respectively (see **Table 5** for MSI rates).

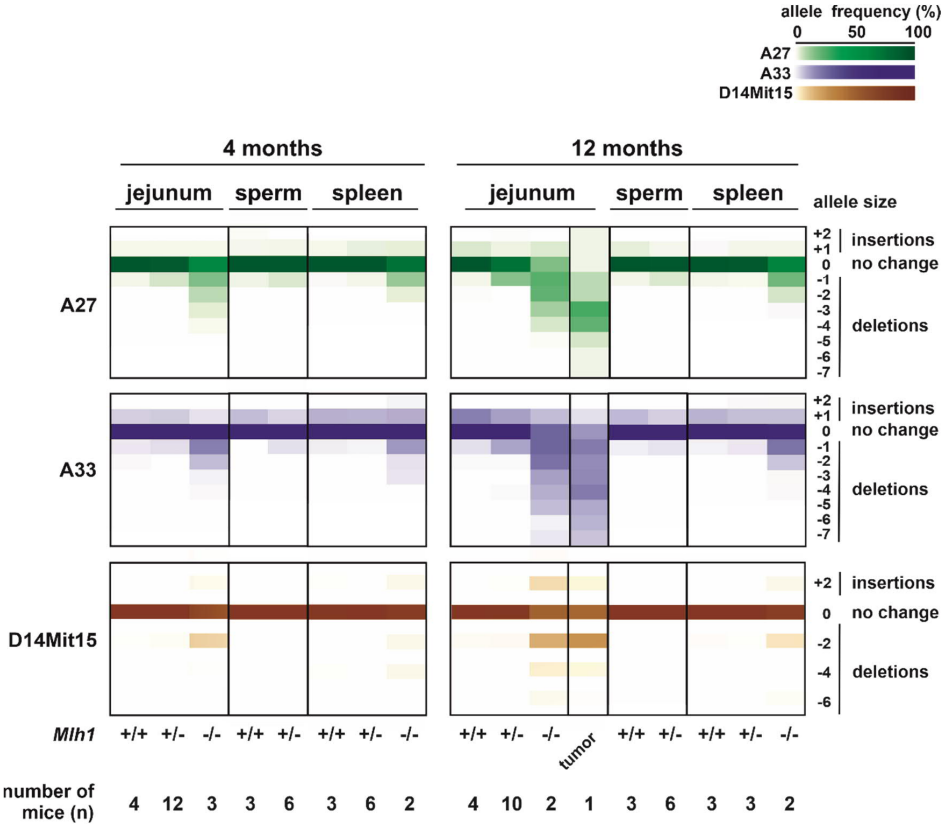


Figure 6. Heat-map of allele frequency (%) at microsatellite loci A27, A33 and D14Mit15. Heat-map shows gradual increase in size-shift and in frequencies of deletion mutants with decreasing *Mlh1* gene dosage and with age, severity of which associates with proliferative capacity of tissues. Deletions further increased in *Mlh1*$^{-/-}$ intestinal tumor. Insertions on the other hand were almost exclusively single repeat unit size-shifts, irrespective of tissue or of genotype. The 4-month old *Mlh1* $^{+/-}$ outlier mouse (see section 4.1.1) was excluded from the heat map (separate heat map for this mouse is shown in **Supplementary figure 3 /I**).

Table 4. Summary of SM-PCR based MSI analysis results at mononucleotide repeats. Shown is the MSI(%) at A37 and A33 ± SD. Data for 4-month outlier mouse is shown in separate column.

	deletions (%)				insertions (%)			
tissues	***Mlh1*** $^{+/+}$	***Mlh1*** $^{+/-}$	**outlier** ***Mlh1*** $^{+/-}$	***Mlh1*** $^{-/-}$	***Mlh1*** $^{+/+}$	***Mlh1*** $^{+/-}$	**outlier** ***Mlh1*** $^{+/-}$	***Mlh1*** $^{-/-}$
4 months								
jejunum	2.4 ± 0.5	5.5 ± 1.7	32.8	50 ± 2.5	3.9 ± 0.9	4.2 ± 2.1	1.9	3.7 ± 0.7
spleen	1.8 ± 0.3	1.8 ± 0.6	4.9	26 ± 0.9	6.6 ± 1	5.9 ± 2.1	3.4	8.4 ± 0.7
sperm	1.7 ± 0.7	4.5 ± 0.5	7.9	NA	6 ± 0.6	4 ± 1.1	4.9	NA
12 months								
jejunum	1.7 ± 0.2	8.1 ± 2.2		99 ± 5.8	6.3 ± 0.2	3.5 ± 1.3		1.9 ± 0
spleen	1.7 ± 1.0	1.8 ± 0.7		41 ± 5.8	5.2 ± 0.8	5.6 ± 1.2		5.5 ± 0.9
sperm	2.6 ± 0.5	5.9 ± 0.7		NA	6 ± 0.9	4.2 ± 0.7		NA

Table 5. Summary of SM-PCR based MSI analysis results at dinucleotide repeat D14Mit15. Presented is average MSI(%) ± SD. Data for 4-month outlier mouse is shown in separate column.

	deletions (%)				insertions (%)			
tissues	***Mlh1*** $^{+/+}$	***Mlh1*** $^{+/-}$	**outlier** ***Mlh1*** $^{+/-}$	***Mlh1*** $^{-/-}$	***Mlh1*** $^{+/+}$	***Mlh1*** $^{+/-}$	**outlier** ***Mlh1*** $^{+/-}$	***Mlh1*** $^{-/-}$
4 months								
jejunum	0.2 ± 0.2	0.7 ± 1.5	3.9	11.2 ± 1.4	0 ± 0	0.2 ± 0.2	0	2.8 ± 0.8
spleen	0.2 ± 0.3	0 ± 0	0	4.7 ± 1.3	0.3 ± 0.5	0 ± 0	0	2.3 ± 1.1
sperm	0 ± 0	0 ± 0	0	NA	0 ± 0	0.2 ± 0.3	0	NA
12 months								
jejunum	0.6 ± 0.4	2.3 ± 2.5		24.6 ± 0.2	0 ± 0	0.1 ± 0.1		3.8 ± 0.9
spleen	0 ± 0	0.4 ± 0.6		6.6 ± 1.8	0 ± 0	0 ± 0		2.7± 0.3
sperm	0.2 ± 0.2	0.1 ± 0.2		NA	0 ± 0	0.1 ± 0.2		NA

4.2. *Mlh1*$^{+/-}$ mice show sporadic decrease in MLH1 expression levels in small intestine but not in spleen (Study I)

Mlh1$^{+/-}$ small intestine (including outlier mouse) showed higher inter-individual variation in deletion frequencies compared to *Mlh1*$^{+/-}$ spleens. To assess whether this inter-individual variation was due to variable *Mlh1* expression levels, we quantified *Mlh1* mRNA and MLH1 protein using qPCR and western blot, respectively. Age-matched spleens which showed minimal MSI in *Mlh1*$^{+/-}$ mice were used as control tissue. Wildtype and *Mlh1* knockout tissues were used as positive and negative controls for MLH1 expression, respectively. The qPCR and western blot experiment were performed in 4- and 12-month time points. We also assessed *Mlh1* expression variation in jejunum of 4-month old *Mlh1*$^{+/-}$ mice by immunohistochemistry (IHC). Further, *Mlh1* mRNA expression was assayed in additional somatic tissues (namely: brain, kidney and liver) for a broader view of tissue-specific differences in *Mlh1* expression.

Mlh1$^{+/-}$ spleens showed the expected, approximately 50% level of *Mlh1*$^{+/+}$ expression both at the transcript and protein levels at both 4- and 12-month time points (**Figure 2A/I and Figure 2B/I**). In contrast, *Mlh1*$^{+/-}$ small intestine showed lower than expected *Mlh1* mRNA and MLH1 protein expression compared to *Mlh1*$^{+/+}$ expression (**Figure 2A/I and Figure 2B/I**). At 4 months, compared to the wildtype expression, the average *Mlh1* mRNA and *Mlh1* protein expression in the *Mlh1*$^{+/-}$ small intestine were 26% (range: 8%-46%) and 24% (range: 0.4%-53%), respectively, and at 12 months, 15% (range: 9%-23%) and 17% (range: 0.2%-29%), respectively (**Figure 2A/I and Figure 2B/I**). *Mlh1* expression decreased in *Mlh1*$^{+/-}$ small intestine with age (from 4- to 12-month time point), however, the decrease was not statistically significant. Further, we assessed *Mlh1* mRNA expression in small intestine of 1-month old *Mlh1*$^{+/-}$ mice. Compared to the 1-month time point, *Mlh1*$^{+/-}$ small intestines at 12-month time point showed significant decrease in *Mlh1* mRNA levels ($p=0.0283$), while *Mlh1*$^{+/-}$ spleens

at the 1-month time point (like 4- and 12- month time point) showed the expected approximately 50% of wildtype *Mlh1* mRNA levels (**Supplementary figure 4/I**). In addition to *Mlh1*$^{+/-}$ spleen and *Mlh1*$^{+/-}$ small intestine, we quantified *Mlh1* mRNA expression also in brain, kidney and liver of 4-month old *Mlh1*$^{+/-}$ mice. All three of the assayed *Mlh1*$^{+/-}$ tissues showed the expected *Mlh1* mRNA expression (60%, 56% and 56% compared to *Mlh1*$^{+/+}$ tissues, respectively) (**Supplementary figure 5/I**).

Based on the MLH1 protein expression levels in the small intestine, *Mlh1*$^{+/-}$ mice could be divided into two groups, from here onward termed as *Mlh1*$^{+/-}$ MLH1^{normal} and *Mlh1*$^{+/-}$ MLH1^{low} and defined by comparatively higher MLH1 protein levels (approximately 50% of the wildtype MLH1 expression, range: 18-53% and 19-32% at 4- and 2-month time points, respectively), versus lower MLH1 protein levels (less than 4% at both time points) in small intestine, respectively (see histogram of **Figure 2B/I**). 7 and 5 out of 12 4-month *Mlh1*$^{+/-}$ mice, and 7 and 4 out of 10 12-month *Mlh1*$^{+/-}$ mice assayed were categorized as *Mlh1*$^{+/-}$ MLH1^{normal} and *Mlh1*$^{+/-}$ MLH1^{low} mice, respectively.

To complement protein expression analysis by western blot, we also performed IHC in small intestine of 4-month old mice. Upon automated quantification and scoring of the staining intensities by IHC profiler (a plug-in in ImageJ software), 3 out of 6 (50%) of the *Mlh1*$^{+/-}$ small intestine was scored as MLH1 low-positive, while other 50% of *Mlh1*$^{+/-}$ small intestine was scored as MLH1-positive (**Figure 2C/I**). Further, assessing the MLH1 staining in the crypt and in the villi separately with IHC, we discovered that MLH1 expression decreased in MLH1 low-positive *Mlh1*$^{+/-}$ small intestine both in the crypt and in the villi, and that MLH1-negative cells were exclusive to villi (**Supplementary figure 6/I**).

In brief, with exception of the *Mlh1*$^{+/-}$ small intestine, all *Mlh1*$^{+/-}$ tissues tested (spleen, brain, kidney and liver) showed the expected, approximately 50% reduced *Mlh1* expression levels compared to wildtype. In *Mlh1*$^{+/-}$ small intestine, substantial numbers of *Mlh1*$^{+/-}$ mice showed lower than expected levels of MLH1 expression, some with undetectable levels of MLH1.

4.3. Depletion of MLH1 in small intestine of *Mlh1*$^{+/-}$ mice associates with soma-wide *Mlh1* promoter methylation (Study I-II)

To investigate possible underlying causes of sporadic tissue-specific MLH1 depletion in *Mlh1*$^{+/-}$ jejunum, we tested for genetic (LOH) and epigenetic (*Mlh1* promoter methylation) mechanisms known to effect the MLH1 expression using RFLP and MSP assay, respectively.

Only one (the 4-month old outlier *Mlh1*$^{+/-}$ mice) out of 22 *Mlh1*$^{+/-}$ mice tested showed LOH in small intestine (**Figure 3B/I**). All other *Mlh1*$^{+/-}$ mice retained the *Mlh1*$^{+/+}$ allele in small intestine (**Figure 3B/I and Supplementary figure 8/I**). Surprisingly, a significant number of *Mlh1*$^{+/-}$ mice (16 of 22) showed *Mlh1* promoter methylation in normal jejunum and spleen. Upon assaying *Mlh1* promoter methylation status in other tissues (namely: brain and sperm), we found *Mlh1* promoter methylated in both the tested tissues (**Supplementary figure 10/I**, **Figure 2/II** and **Supplementary figure 4/II**). All the observed *Mlh1* promoter methylation was partial, i.e. DNA fractions both with and without *Mlh1* promoter methylation were detected. Further, *Mlh1*$^{-/-}$ mice also showed the similar (i.e. partial) *Mlh1* promoter methylation in all tissues analyzed, while none of the tissues tested in *Mlh1*$^{+/+}$ mice showed *Mlh1* promoter methylation (**Figure 3C/I, Supplementary figure 10/I**, **Figure 2/II** and **Supplementary figure 4/II**).

4.4. MSI correlates with decreased MLH1 protein levels in small intestine of *Mlh1*$^{+/-}$ mice (Study I-II)

We also studied the correlation between MLH1 protein levels and deletions (at mononucleotide repeats) in *Mlh1*$^{+/-}$ small intestine and spleens. At the 4-month time point, small intestine of *Mlh1*$^{+/-}$ MLH1^{normal} and *Mlh1*$^{+/-}$ MLH1^{low} mice showed close-to-expected (35%) and very low (2%) average MLH1 protein expression, and displayed 3.9% and 7.2% deletions, respectively (See **Table 6** and **Figure 4/I**). At the 12-month time point, small intestine of *Mlh1*$^{+/-}$ MLH1^{normal} and *Mlh1*$^{+/-}$ MLH1^{low} mice showed lower-than-expected (24%) and very low (1%) average MLH1 protein expression, and exhibited 6.6% and 11.2% deletions, respectively (see **Figure 4/I**). The *Mlh1*$^{+/-}$ outlier mouse had almost no detectable MLH1 protein expression in its jejunum and displayed 32.8% deletions (see **Figure 4/I**). Compared to 4 months, *Mlh1*$^{+/-}$ MLH1^{normal} mice at 12 month time point showed 1.5-fold decrease in MLH1 expression and showed 1.7-fold increase in deletions (see **Figure 4/I**). *Mlh1*$^{+/-}$ MLH1^{low} mice at 12 months showed 1.5-fold higher deletions than at 4-month time point, while the MLH1 expression levels between the two groups were comparable. All *Mlh1*$^{+/-}$ spleen samples (at both time points) showed expected levels (i.e. approximately half of wildtype MLH1 expression), and comparable deletions to *Mlh1*$^{+/+}$ spleen, except in the outlier *Mlh1*$^{+/-}$ mouse which showed a 2-fold increase in deletions in spleen compared to age-matched *Mlh1*$^{+/-}$ spleens, despite having normal protein expression level (**Supplementary Figure 12/I**).

We also performed linear regression analysis to study the correlation of MLH1 protein expression and MSI rates in the small intestine. The *Mlh1*$^{+/-}$ small intestine were clustered into two distinct groups. MLH1 protein expression levels inversely correlated with deletion rates at both the time points (R^2 = 0.5, p= 0.001 and R^2 = 0.46, p= 0.01 for 4- and 12- month time point, respectively) (**Supplementary Figure 11/I**).

Table 6. Comparison of MLH1 expression levels and deletion frequencies in small intestine of MLH1^{normal} and MLH1^{low} *Mlh1*$^{+/-}$ mice.

	MLH1^{normal}		MLH1^{low}	
	average (%)	**range (%)**	**average (%)**	**range (%)**
	4 months			
MLH1 protein level	35	27 - 53	2	0 - 4
deletions	3.9	2.8 - 4.4	7.2	5.6 - 8.8
	12 months			
MLH1 protein level	24	18 - 29	1	1 - 2
deletions	6.6	5.7 - 7.7	11.2	9.8 - 12.7

We grouped the *Mlh1*$^{+/-}$ sperm samples based on the *Mlh1*$^{+/-}$ MLH1^{normal} and *Mlh1*$^{+/-}$ MLH1^{low} classification. Unlike in the small intestine, there was no substantial difference in sperm DNA deletions between the two *Mlh1*$^{+/-}$ sub groups. However, *Mlh1*$^{+/-}$ mice with soma-wide *Mlh1* promoter methylation in general showed higher deletion rate in sperm (**Figure 7**).

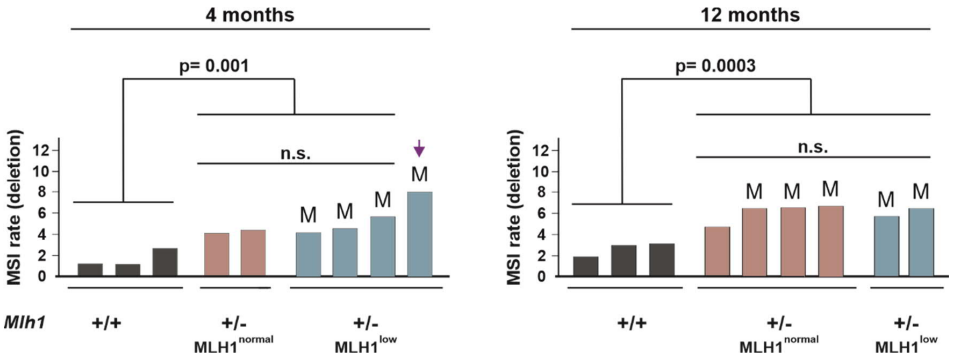


Figure 7. Sperm of MLH1^{normal} and MLH1^{low} *Mlh1*$^{+/-}$ mice show comparable deletions at mononucleotide repeats. "M" indicates sperm with *Mlh1* promoter methylation. Arrow indicates the outlier *Mlh1*$^{+/-}$ mouse.

5. DISCUSSION

Individuals with an inherited mutant MMR allele are born with one functional copy of the MMR gene and their tissues are thought to be microsatellite stable. LS associated CRCs are MMR-deficient and show MSI-high phenotype, which led to the notion that MSI occurs only after complete loss of MMR (i.e. after the "second hit" of the MMR genes) (Aaltonen et al., 1993; Hemminki et al., 1994). Evidence of LS-associated adenomas retaining (some) MMR protein expression suggests that the second hit may occur later in the multi-step tumorigenesis than previously anticipated, however these adenomas were classified as microsatellite stable (Valo et al., 2015). Other clues suggest that MSI may occur (rather early) in presence of MMR. For example, low-level MSI in seen in peripheral blood leukocytes (PBLs) of LS individuals (Alazzouzi et al., 2005; M. I. Coolbaugh-Murphy et al., 2010), MSI in normal colon mucosa of LS individuals with minimal MMR expression (Parsons et al., 1995)). We showed that if sufficiently sensitive assays are used, MSI is detectable pre-neoplastic normal tissues *in vivo* in the presence of MMR. Our approach of assaying MSI at highly unstable microsatellites in single DNA-molecule level detects MSI lower than 1%, while standard MSI assays (which use bulk DNA) have a sensitivity limit of 20-25 % MSI (M. Coolbaugh-Murphy et al., 2004). Our highly sensitive SM-PCR assay can quantify baseline levels of MSI even in *Mlh1*$^{+/+}$ tissues. We showed that MSI, particularity deletions at microsatellites is dependent on MLH1 protein levels in somatic tissues. *Mlh1*$^{-/-}$ tissues showed high-level of MSI, while *Mlh1*$^{+/+}$ tissues were microsatellite stable or showed base-line levels of deletions. In highly proliferating *Mlh1*$^{+/-}$ small intestine, we showed a sporadic MLH1 depletion and elevated MSI, while no such MLH1 level aberration was observed in *Mlh1*$^{+/-}$ spleens and was microsatellite stable. In *Mlh1*$^{+/-}$ small intestine, firstly, we showed that *Mlh1*$^{+/-}$ mice with close to expected MMR protein levels (i.e. approximately 50% of the wildtype expression) exhibit MSI in the small intestine, providing evidence of tissue-specific *Mlh1* haploinsufficiency. Secondly, MSI further increases with sporadic decrease in MLH1 protein levels in the *Mlh1*$^{+/-}$ small intestine. In *Mlh1*$^{+/-}$ small intestine, we observed a critical threshold of MLH1 protein level (18% compared to wildtype expression) below which MSI

exacerbates. Similar critical threshold and loss of MMR function was reported *in vitro* by Kansikas et. al., they reported loss of MMR function below 23% Mlh1 mRNA expression level (Kansikas, Kasela, Kantelinen, & Nystrom, 2014). Our results show that small intestine in *Mlh1*$^{+/-}$ mice is particularly prone to aberrant MLH1 depletion, and that normal small intestine of *Mlh1*$^{+/-}$ mice show MSI. Likely similar pre-neoplastic cellular aberrations occur in the GI tract of LS individuals, instigating GI tract tumorigenesis.

We did not observed any tumor in the small intestine of *Mlh1*$^{+/-}$ mice. What we have reported here (MSI readouts at non-exonic microsatellites) likely represents early-stage pre-tumorigenic genomic instability. MSI *per se* does not necessary lead to tumorigenesis (Prolla et al., 1998). For tumorigenesis to initiate, MSI should occur in exonic microsatellites (i.e. in the MSI-target genes) (Duval & Hamelin, 2002; Vilkki et al., 2002). MSI in exonic microsatellites could alter gene function and instigate tumorigenesis (Duval & Hamelin, 2002). The exons are better protected by the MMR (Frigola et al., 2017) than other genic regions, thus, exons are unlikely to accumulate early mutations which could be catastrophic to the cells. We focused our study on non-exonic microsatellites which are less protected by MMR due to differential MMR (Frigola et al., 2017), making them more sensitive indicator of MMR defects. This gave us the advantage of detecting early-stage pre-tumorigenic MSI. Although we did not observe any tumor in *Mlh1*$^{+/-}$ small intestines, we uncovered large inter-individual variation in MSI in normal (non-neoplastic) small intestine of *Mlh1*$^{+/-}$ mice, ranging from MSI comparable to *Mlh1*$^{+/+}$ to that in *Mlh1*$^{-/-}$ small intestine. We also observed an age-dependent increase in MSI levels in *Mlh1*$^{+/-}$ small intestine. Moreover, an appreciable fraction (42%) of young (4-months) *Mlh1*$^{+/-}$ mice showed MSI levels in their pre-neoplastic small intestine comparable to old (12-months) *Mlh1*$^{+/-}$ mice. Similar observations i.e. some young individuals showing elevated MSI were reported in peripheral blood leukocyte DNA of LS patients (M. I. Coolbaugh-Murphy et al., 2010). Further, one 4-month old *Mlh1*$^{+/-}$ mice showed high MSI in its (normal and tumor-free) small intestine, comparable to MSI in age-matched *Mlh1*$^{-/-}$ small intestine and higher

than MSI observed in 12-month *Mlh1*$^{+/-}$ small intestine. If this mouse was aged more it would have likely accumulated mutations in the MSI-target genes triggering tumorigenesis in the GI tract. In addition, 30% of *Mlh1*$^{+/-}$ mice died in our colony before they were used in experiments but were not examined post-mortem (see Kaplan-meier survival curve in **Supplementary figure 13/I**). Likely some, if not all, of these *Mlh1*$^{+/-}$ mice died because of MMR-associated tumors. Overall, our results show that *Mlh1*$^{+/-}$ small intestine is susceptible to tissue-specific pre-neoplastic MSI *in vivo*, accumulation of which starts already at a young age and varies substantially between individuals.

We observed MSI in germline cells (sperm) of *Mlh1*$^{+/-}$ mice, and MSI increased with age. A mature sperm cell is a product of multiple cell divisions (a spermatogonial stem cell (SSC) undergoes up to 12 cell divisions to produce a mature sperm). In such a highly proliferating system the likelihood of accumulating replication error increases with each round of DNA replication, thus MSI in *Mlh1*$^{+/-}$ sperm is not surprising. Comparing MSI in sperm to that of the somatic tissues (jejunum and spleen) of the same *Mlh1*$^{+/-}$ mice, all *Mlh1*$^{+/-}$ mice showed higher MSI in sperm compared to spleen. Compared to small intestine, 44% of *Mlh1*$^{+/-}$ mice showed similar levels of MSI in their sperm, while in the remaining *Mlh1*$^{+/-}$ mice (56%) small intestine showed higher MSI than sperm (**Figure 8**). A likely explanation of sporadic increase in MSI in small intestine compared to sperm of *Mlh1*$^{+/-}$ mice is our observation of sporadic decrease in MLH1 levels in small intestine. Although we did not measure MLH1 levels in testis, since we did not observe sporadic increase in MSI levels in *Mlh1*$^{+/-}$ sperm, equitable assumption is that MLH1 levels in *Mlh1*$^{+/-}$ testis is normal (i.e. approximately 50% of the wildtype). Under this assumption *Mlh1*$^{+/-}$ testis are also haploinsufficient as *Mlh1*$^{+/-}$ sperm showed increased MSI compared to *Mlh1*$^{+/+}$ sperm despite possibly having normal MLH1 levels in testis, which needs further investigation. Further, the age when intestine stem cells (ISCs) and SSCs start to proliferate is different. SSCs are in dormant stage (i.e. do not proliferate) until puberty (approximately 4-5 week in male mice, and 15 years in human), but ISCs start to proliferate between E16.5 and postnatal day 7 in mice (Barker, 2014) and already in embryonic stage (week 9-10 of

gestation in human (Drozdowski, Clandinin, & Thomson, 2010). Given that ISCs start to proliferate much earlier in life, and that they have a very high proliferation rate (300 million new epithelial cells are generated every day) (Barker, 2014), by the time SSCs start to proliferate, cells in the small intestine have had numerous opportunities to accumulate mutations. Likely explaining the sporadic increase in MSI in *Mlh1*$^{+/-}$ small intestine compared to *Mlh1*$^{+/-}$ sperm. Unlike *Mlh1*$^{+/-}$ small intestine where we observed substantial inter-individual variation in MSI, *Mlh1*$^{+/-}$ sperm did not show such drastic inter-individual variation (expect the outlier *Mlh1*$^{+/-}$ mouse). Overall, our results indicate that in the *Mlh1* heterozygous condition, the MMR repair activity is compromised to the extent that it is unable to maintain wildtype-level genome stability in highly proliferating somatic and germline cells.

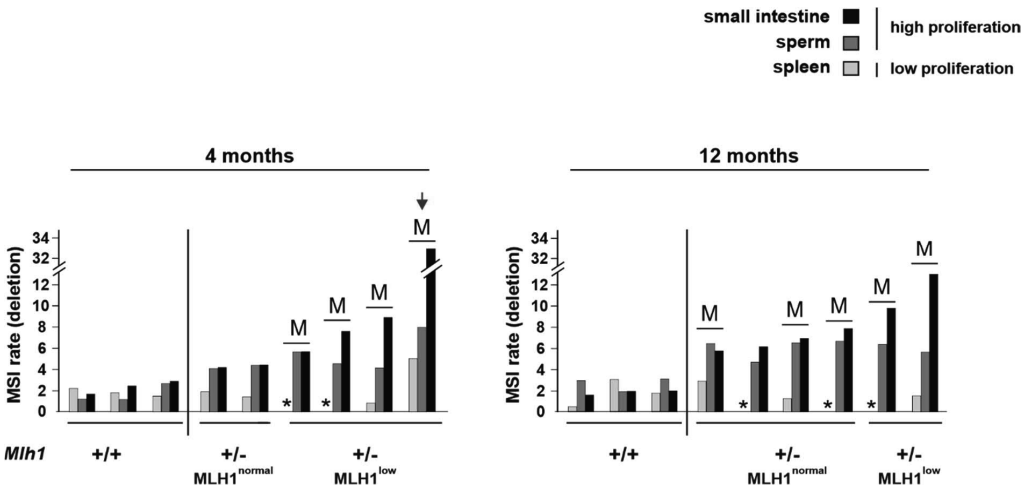


Figure 8. Comparison of tissue-specific deletions at mononucleotide repeats. Highly proliferative small intestine and sperm show higher deletions at mononucleotide repeats than the low proliferating spleen. Further, small intestine show sporadic increase in the deletions in some *Mlh1*$^{+/-}$ mice with soma-wide *Mlh1* promoter methylation (indicated by "M"). Arrow head indicates the outlier mice which showed high MSI in small intestine. Asterisk indicates no data.

We observed mononucleotide repeats are more unstable than the dinucleotide repeat, similar observations has been reported by others (Bacher, Abdel Megid, Kent-First, & Halberg, 2005; Kabbarah et al., 2003). Our result highlights the importance of selecting appropriate microsatellite loci especially for detection subtle MSI phenotype. Interestingly, we found different dynamics of insertions versus deletions at the mononucleotide repeats. Firstly, we observed an insertional bias (particularly at A33) in wildtype tissues (both in germline and somatic cells) while deletions were negligible. Secondly, in highly proliferating tissues, deletions tended to increase with deceasing *Mlh1* gene dosage while insertions showed the opposite trend; low proliferating spleen, on the other hand, did not show such *Mlh1*-dosage dependent fluctuations in MSI. Further, while we observed almost exclusively single repeat unit (i.e. 1bp) insertions irrelevant of tissue type or gene dosage, deletion-allele size increased with decreasing *Mlh1* dosage in proliferating tissues. Like in normal tissues, the *Mlh1*-deficient intestinal tumor also displayed a high level of deletions and almost no detectable insertions.

As the endogenous source of insertions and deletions at microsatellites is the same (i.e. replication errors made by the faulty DNA polymerase), one would expect the likelihood of insertions and deletions at microsatellites to be similar. However, we saw clear opposite dynamics of insertions versus deletions at the mononucleotide repeats in proliferating tissues. DNA loop formation due to replication slippage in the nascent or template strand gives rises to insertion or deletion at microsatellites, respectively (see **Figure 3**). Post-replication these indels should be repaired by MMR. Our observation of insertional burden but minimal deletion in wildtype tissues implies formation of DNA loops primarily occur on the nascent strand and/or less efficient repair of insertions by MMR. Lower insertion rates in *Mlh1* hetero- and homozygous tissues can be likely explained by deletion events erasing many of the insertions that occurred in the earlier cell division converting the 1bp insertion mutant alleles into wild-type alleles.

We detected *Mlh1* promoter methylation in $Mlh1^{+/-}$ and $Mlh1^{-/-}$ tissues originating from all embryonic layers: endoderm (jejunum), ectoderm (brain) and mesoderm (spleen), which indicates that *Mlh1* promoter methylation has occurred in the germline or in the early-post-zygotic stage as suggested by others who observed the same phenotype in multiple human tissues (Niessen et al., 2009). Also, we observed *Mlh1* promoter

methylation in *Mlh1*$^{+/-}$ sperm. Ours is the first observation of soma-wide *Mlh1* promoter methylation phenotype in a MMR mouse model. In human, this phenotype has been reported in LS (suspected) patients (Suter et al., 2004). Despite observing *Mlh1* promoter methylation in multiple tissues of *Mlh1*$^{+/-}$ mice, we observed MLH1 depletion only in the *Mlh1*$^{+/-}$ small intestine, thus *Mlh1* promoter methylation cannot by itself account to lower MLH1 levels in small intestine. Most likely there are other regulatory mechanism/s unknown so far that govern the MMR protein levels in the GI tract which needs further investigation.

LS (suspected) patients and LS patients both have the same tumor spectra, however LS (suspected) patients have relatively early onset of tumor compared to the LS patients. In our study, elevated MSI in *Mlh1*$^{+/-}$ proliferating tissues associated with *Mlh1* promoter methylation. It is likely that LS (suspected) patients also have relatively earlier accumulation of MSI compared to LS patients, leading to early appearance of tumor compared to the LS patients. In addition, sperm in LS (suspected) patients are also reported to exhibit *Mlh1* promoter methylation (Suter et al., 2004). We show that *Mlh1* promoter methylation in *Mlh1*$^{+/-}$ sperm associates with MSI in the *Mlh1*$^{+/-}$ sperm. Human studies have shown a high association of single nucleotide polymorphism (SNP) in *Mlh1* promoter and the occurrence of *Mlh1* promoter methylation (Suter et al., 2004). Our study mice were a cross of C57BL/6J and 129S1/SvImJ mouse strains. These mice have SNPs in and around the core promoter, and it is possible that there is preferential methylation of *Mlh1* promoter in one of the two alleles (either C57BL/6J or 129S1/SvImJ), which needs further investigation. Our observation of constitutive *Mlh1* promoter methylation in *Mlh1*$^{-/-}$ mice (mice of 129S1/SvImJ background), and no *Mlh1* promoter methylation in the tissues analyzed in the *Mlh1*$^{+/+}$ mice (mice of C57BL/6J background) indicates that *Mlh1*$^{-/-}$ allele is preferentially methylated. Further, LOH was observed in only one (of 22) *Mlh1*$^{+/-}$ mice, indicating that *Mlh1* promoter methylation is an earlier and more common event than LOH in multi-step tumorigenesis at least in this mouse model. Overall we found that *Mlh1*$^{+/-}$ mice exhibit MSI in proliferating normal (non-neoplastic) somatic tissue (small intestine) and in germline cells. Further, we observed age independent sporadic MLH1 depletion in the

normal small intestine and elevated MSI in the normal small intestine of these mice. *Mlh1* promoter methylation only affected MLH1 protein expression in small intestine of the *Mlh1*$^{+/-}$ mice and not the other tissues assayed, as such, *Mlh1* promoter methylation (at least not alone) cannot explain MLH1 protein depletion in the small intestine, however the observation of *Mlh1* promoter methylation, MLH1 protein level depletion, and MSI in *Mlh1*$^{+/-}$ small intestine indicate these observation are very-early pre-tumorigenic event.

6. CONCLUSION AND FUTURE PROSPECTIVE

My thesis presents a comprehensive study on the effect of MMR heterozygosity on somatic and germline genomic instability. Further, my thesis elucidates the correlation between tissue-specific genomic instability and MMR expression levels in somatic tissues with variable (i.e. high and low) proliferation rates.

Using a highly sensitive single-molecule PCR based MSI assay, we were able to detect MSI in *Mlh1*$^{+/-}$ tissues; we showed that highly proliferating *Mlh1*$^{+/-}$ small intestine and *Mlh1*$^{+/-}$ sperm display MSI which increases with age, while the low proliferating spleen was microsatellite stable. Further, we show that the microsatellites (in particular mononucleotide repeats) are more prone to deletions. In addition, we show that soma-wide *Mlh1* promoter methylation is common in *Mlh1*$^{+/-}$ mice and associates with observed MSI in proliferating tissues.

Further, investigating the relationship between tissue-specific MLH1 expression and MSI in somatic tissues, we show that specifically the *Mlh1*$^{+/-}$ small intestine is vulnerable to elevated MSI, which seems to depend on a critical threshold of MLH1 protein level, and is age-independent. Firstly, we showed that the *Mlh1*$^{+/-}$ small intestine expressing close to expected MLH1 levels (i.e. approximately 50% of wildtype) display MSI, implying MMR haploinsufficiency in normal, tumor-free *Mlh1*$^{+/-}$ intestine. Secondly, we show that when MLH1 expression reaches a critical threshold (below approximately 20% of *Mlh1*$^{+/+}$ expression), MSI (particularly deletions) substantially increases in the *Mlh1*$^{+/-}$ intestine.

Overall, we showed that the aberrant molecular phenotypes displayed by MMR-associated tumor (namely: MSI, *Mlh1* promoter methylation and MMR-deficiency) is seen, though at relatively low levels, in normal, tumor-free *Mlh1*$^{+/-}$ intestine at a young age. Thus, these phenotypes, previously considered as cancer-associated events occurs very early as a tumor precursors in a tissue-specific manner, which likely explains why the GI tract of LS individuals are particularly prone to MSI-high tumors.

7. ACKNOWLEDGEMENTS

The studies compiled in this thesis were conducted in Genome Stability group (Kauppi lab). The PhD journey took 1/5 years of my current life; pretty big chunk of my life. I am very thankful to all of my colleagues, friends, well-wishers, academics and my family for your love and support in this roller-coaster ride.

Firstly, I want to thank my guru in research Associate Prof. Liisa Kauppi. Thank you for your guidance and your support through-out my PhD journey. You have been the pillar in up-building my research career. Liisa took me in her lab as a summer intern some 8 years back, significant metamorphosis from then to a master's thesis student and to this moment when I'm writing final words of my PhD thesis, thank you for all the opportunities, I am forever grateful!

I want to thank all the academics involved in my PhD journey. Thank you Associate Prof. Francesca Cole for agreeing to be my opponent, looking forward to the defence! Thank you my thesis pre-examiners Adjunct Prof. Keijo Viiri and Prof. Yuri Dubrova for examining my thesis and for very motivating- and positive- statements. Thankyou Prof. Juha Partanen for agreeing to be my custos. Thank you Prof. Kari Keinänen, my responsible professor, for helping me with all the practical matters regarding PhD dissertation submission and acceptance. My thesis committee Dr. Saara Ollila and Dr. Riikka Kivelä, thank you for guiding me to the end of the tunnel.

I would like to thank Biomedicum Helsinki Foundation and Lasten syöpäsäätiö Väre for personal grants during my PhD studies. Thank you FIMM genomics core (especially Suvi) and Immuno-histochemisty core (Reeta and Helena) for running all the PCR plates and for IHC help, respectively.

I would like to thank past and present members of Kauppi lab with whom I got to share my time in lab and out. Thank you for all the memories Maarit, Dr. Juha, Dr. Denis, Dr. Manu, Dr. Nanna, Dr. Linlin, Dr. Imrul, Dr. Barun, Elli, Sanna, Minna, Taina, Sundarya, Sonja, Susanna, Venessa, Henrik and Venkatram. Special thanks to Maarit, Denis, Elli and Minna who have helped me a lot in my projects (Team MMR, Woohoo!). I remember Maarit helping me pour my first agarose gel in lab, we shared exciting times with my then favorite gene PRDM9. Thankyou Maarit for being such a caring person. Thankyou Denis for your intellectual input in my early PhD days. Elli, thank you for co-

authoring in my research paper and all the fun and goofy times. Minna, I cannot even imagine my PhD journey without you, your contribution to my projects is tremendous. Thank you for being such an efficient co-worker, a good and caring friend, and for co-authoring in both my research papers.

I also want to thank my next-door-lab neighbors, my dear protein people: Dr. Erika, Elina, Ani, Jonas, Kate, Bia, Nastia for all the protein related help and consultations. I never felt alone working late hours or in the weekends because of you all. Thank you for the memorable times both in and outside of lab!

Thank you my academic social group for all the fun times! My ILS family, all the memories of organizing and participating in the events together is very dear to me. A shout out to the past and present ILS council members Beh, Dr. Alok, Isabel, Larrisa, Iilda, Tuomo, Mirdul, Dr. Markku, Heini, Dr. Geri, Dr. Jarno, Siggi, Selma, Dr. Kornelia, Dr. Maarja and Johanna. Sharing all the PhD frustrations was also part of the game, “share your PhD frustration sessions” were indeed very consoling. I also want to thank the TedxUH friends, rowing team and friends in Biomedicum for keeping me active. Special thanks to Dr. Chiara, thank you for all the hugs, memories and fun times.

Music has always been an integral part of my life, my escape from reality. Thanks to the legendary academic band “The UnderDocs”, I had tons of fun time with you playing as a band on stage and off stage on weekly rehearsals during my PhD. I would like to thank past/present band members Prof. Juha, Dr. Mike, Giorgio, Dr. Heidi, Fon and Tapio for amazingly good times, keep on rocking! Dr. Julia, Hans, Prima and Mikko, it was short but a lovely time playing as a band together, keep on rocking! Shout out to Markko, Kasper, Mekhail, Jaska and Ricardo, members of my current band BlackLight Orchestra, thank you for the weekly music therapy during my PhD, therapy mantra: keep it simple, heavy, groovy and loud!

August 27, 2012 was when I came to Finland, thank you all my Nepalese bundle for making me feel home away from home. Nepalese bundle of joy in Biomedicum: Dr. Shishir, Dr. Sawan, Dr. Prson, Dr. Abishesk, Dr. Bhagwan, Khusbu, Sweta, and Puspa, my Aalto bundle of joy: Prasamsha, Dr. Gautam, Jyoti, Dr. Prajwol, Shristi, Pratik, Pratistha, Dr. Bishal, Dr. Neha, Dr. Subash and Rashmi, and Nepalese bundle of joy in Helsinki: Dr. Anirudra, Bina, Biplu, Bideep, Prabin, Rabina, Ravi, Chhabi, Arun, Subash, Dr. Mamta, Pramod, Dr. Sudeep, Dr. Sanju, Sadin and Elina thank you all for

friendship, good food and good times. Special mention to my swimming and cycling bundle Niwas, Bibhor and Abina, thank you for keeping me fit!

Katja you are my very first Finnish friend, thank you for being there since day 1 of my Finland days!

Team Resistomap, my new team! Thankyou for taking me in so swiftly and motivating and supporting me in final leg of my PhD! Thank you Windi, William, Jesse, Tilkku, Alma, Lotta, Irwin, Aloy and Mattias for such a welcoming gesture! Looking forward to the Resistomap journey.

Shout out to my IsotopeZ bundle, my brothers from another mother: Pravin, Abi, Sahem, Saugat, Sagun, Dr. Anan, Dr. Saurav, Pallav and Sagar. Thank you for your everlasting friendship, and for always being fascinated with my work, believing in me, and motivating me a lot in my PhD days! Kids doing crazy experiments with life, science and music, doing interesting (successful/failed) science experiments in the school days, building a voice-controlled robot ("The Hexapod") which actually worked in college days. You guys have helped me a lot to develop and polish my scientific curiosities. Well, life itself is an experiment, happy we are onboard this journey together.

I would like to thank my dearest friend Anjani for always being there for me in my good times and in bad. The comfort and joy of having you by my side has always helped me overcome all the challenges in life. Thank you for being there in my PhD journey, thank you for encouraging me, thank you for giving me hope, thank you for being my support system!

Last but not the least, I want to thank my lovely family: mamu, daddy, brother (dai), sister-in-law (bhauju), niece (nanu) and nephew (babu) for all the love and support. Thank you mamu and daddy for believing in me and for unconditional love. For past 9 year when I was abroad (PhD taking significant chunk of these years), every single morning my mamu and daddy skyped me. Every day they gave me their best wishes and their blessings. Every single day mamu told me to be/do my best and success will follow. Well, Mamu and daddy, I dedicate my PhD to both of you.

-Kul Shanker Shrestha, Helsinki 2021

8. REFERENCES

Aaltonen, L. A., Peltomaki, P., Leach, F. S., Sistonen, P., Pylkkanen, L., Mecklin, J. P., . . . et al. (1993). Clues to the pathogenesis of familial colorectal cancer. *Science, 260*(5109), 812-816.

Aaltonen, L. A., Peltomaki, P., Mecklin, J. P., Jarvinen, H., Jass, J. R., Green, J. S., . . . et al. (1994). Replication errors in benign and malignant tumors from hereditary nonpolyposis colorectal cancer patients. *Cancer Res, 54*(7), 1645-1648.

Alazzouzi, H., Domingo, E., Gonzalez, S., Blanco, I., Armengol, M., Espin, E., . . . Schwartz, S., Jr. (2005). Low levels of microsatellite instability characterize MLH1 and MSH2 HNPCC carriers before tumor diagnosis. *Hum Mol Genet, 14*(2), 235-239. doi:10.1093/hmg/ddi021

Antequera, F., & Bird, A. (1993). Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci U S A, 90*(24), 11995-11999. doi:10.1073/pnas.90.24.11995

Bacher, J. W., Abdel Megid, W. M., Kent-First, M. G., & Halberg, R. B. (2005). Use of mononucleotide repeat markers for detection of microsatellite instability in mouse tumors. *Mol Carcinog, 44*(4), 285-292. doi:10.1002/mc.20146

Baker, S. M., Bronner, C. E., Zhang, L., Plug, A. W., Robatzek, M., Warren, G., . . . Liskay, R. M. (1995). Male mice defective in the DNA mismatch repair gene PMS2 exhibit abnormal chromosome synapsis in meiosis. *Cell, 82*(2), 309-319.

Baker, S. M., Plug, A. W., Prolla, T. A., Bronner, C. E., Harris, A. C., Yao, X., . . . Liskay, R. M. (1996). Involvement of mouse Mlh1 in DNA mismatch repair and meiotic crossing over. *Nat Genet, 13*(3), 336-342. doi:10.1038/ng0796-336

Barker, N. (2014). Adult intestinal stem cells: critical drivers of epithelial homeostasis and regeneration. *Nat Rev Mol Cell Biol, 15*(1), 19-33. doi:10.1038/nrm3721

Beal, M. A., Rowan-Carroll, A., Campbell, C., Williams, A., Somers, C. M., Marchetti, F., & Yauk, C. L. (2015). Single-molecule PCR analysis of an unstable microsatellite for detecting mutations in sperm of mice exposed to chemical mutagens. *Mutat Res, 775*, 26-32. doi:10.1016/j.mrfmmm.2015.03.010

Bois, P., Stead, J. D., Bakshi, S., Williamson, J., Neumann, R., Moghadaszadeh, B., & Jeffreys, A. J. (1998). Isolation and characterization of mouse minisatellites. *Genomics, 50*(3), 317-330. doi:10.1006/geno.1998.5329

Boland, C. R., & Goel, A. (2010). Microsatellite instability in colorectal cancer. *Gastroenterology, 138*(6), 2073-2087 e2073. doi:10.1053/j.gastro.2009.12.064

Bonadona, V., Bonaiti, B., Olschwang, S., Grandjouan, S., Huiart, L., Longy, M., . . . French Cancer Genetics, N. (2011). Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. *JAMA, 305*(22), 2304-2310. doi:10.1001/jama.2011.743

Burt, R. W. (2000). Colon cancer screening. *Gastroenterology, 119*(3), 837-853.

Chen, P. C., Dudley, S., Hagen, W., Dizon, D., Paxton, L., Reichow, D., . . . Lipkin, S. M. (2005). Contributions by MutL homologues Mlh3 and Pms2 to DNA mismatch repair and tumor suppression in the mouse. *Cancer Res, 65*(19), 8662-8670. doi:10.1158/0008-5472.CAN-05-0742

Coolbaugh-Murphy, M., Maleki, A., Ramagli, L., Frazier, M., Lichtiger, B., Monckton, D. G., . . . Brown, B. W. (2004). Estimating mutant microsatellite allele frequencies in somatic cells by small-pool PCR. *Genomics, 84*(2), 419-430. doi:10.1016/j.ygeno.2004.03.007

Coolbaugh-Murphy, M. I., Xu, J. P., Ramagli, L. S., Ramagli, B. C., Brown, B. W., Lynch, P. M., . . . Siciliano, M. J. (2010). Microsatellite instability in the peripheral blood leukocytes of HNPCC patients. *Hum Mutat, 31*(3), 317-324. doi:10.1002/humu.21190

Covarrubias-Pazaran, G., Diaz-Garcia, L., Schlautman, B., Salazar, W., & Zalapa, J. (2016). Fragman: an R package for fragment analysis. *BMC Genet, 17*, 62. doi:10.1186/s12863-016-0365-6

Damaso, E., Canet-Hermida, J., Vargas-Parra, G., Velasco, A., Marin, F., Darder, E., . . . Pineda, M. (2019). Highly sensitive MLH1 methylation analysis in blood identifies a cancer patient with low-level mosaic MLH1 epimutation. *Clin Epigenetics, 11*(1), 171. doi:10.1186/s13148-019-0762-6

de Koning, A. P., Gu, W., Castoe, T. A., Batzer, M. A., & Pollock, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet, 7*(12), e1002384. doi:10.1371/journal.pgen.1002384
de Wind, N., Dekker, M., Berns, A., Radman, M., & te Riele, H. (1995). Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. *Cell, 82*(2), 321-330. doi:10.1016/0092-8674(95)90319-4
de Wind, N., Dekker, M., Claij, N., Jansen, L., van Klink, Y., Radman, M., . . . te Riele, H. (1999). HNPCC-like cancer predisposition in mice through simultaneous loss of Msh3 and Msh6 mismatch-repair protein functions. *Nat Genet, 23*(3), 359-362. doi:10.1038/15544
de Vries, S. S., Baart, E. B., Dekker, M., Siezen, A., de Rooij, D. G., de Boer, P., & te Riele, H. (1999). Mouse MutS-like protein Msh5 is required for proper chromosome synapsis in male and female meiosis. *Genes Dev, 13*(5), 523-531. doi:10.1101/gad.13.5.523
Drake, J. W., Charlesworth, B., Charlesworth, D., & Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics, 148*(4), 1667-1686.
Drozdowski, L. A., Clandinin, T., & Thomson, A. B. (2010). Ontogeny, growth and development of the small intestine: Understanding pediatric gastroenterology. *World J Gastroenterol, 16*(7), 787-799. doi:10.3748/wjg.v16.i7.787
Dumbovic, G., Forcales, S. V., & Perucho, M. (2017). Emerging roles of macrosatellite repeats in genome organization and disease development. *Epigenetics, 12*(7), 515-526. doi:10.1080/15592294.2017.1318235
Duraturo, F., Liccardo, R., De Rosa, M., & Izzo, P. (2019). Genetics, diagnosis and treatment of Lynch syndrome: Old lessons and current challenges. *Oncol Lett, 17*(3), 3048-3054. doi:10.3892/ol.2019.9945
Duval, A., & Hamelin, R. (2002). Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability. *Cancer Res, 62*(9), 2447-2454.
Eckert, K. A., & Hile, S. E. (2009). Every microsatellite is different: Intrinsic DNA features dictate mutagenesis of common microsatellites present in the human genome. *Mol Carcinog, 48*(4), 379-388. doi:10.1002/mc.20499
Edelmann, W., Cohen, P. E., Kane, M., Lau, K., Morrow, B., Bennett, S., . . . Kucherlapati, R. (1996). Meiotic pachytene arrest in MLH1-deficient mice. *Cell, 85*(7), 1125-1134.
Edelmann, W., Cohen, P. E., Kneitz, B., Winand, N., Lia, M., Heyer, J., . . . Kucherlapati, R. (1999). Mammalian MutS homologue 5 is required for chromosome pairing in meiosis. *Nat Genet, 21*(1), 123-127. doi:10.1038/5075
Edelmann, W., Umar, A., Yang, K., Heyer, J., Kucherlapati, M., Lia, M., . . . Kucherlapati, R. (2000). The DNA mismatch repair genes Msh3 and Msh6 cooperate in intestinal tumor suppression. *Cancer Res, 60*(4), 803-807.
Edelmann, W., Yang, K., Kuraguchi, M., Heyer, J., Lia, M., Kneitz, B., . . . Kucherlapati, R. (1999). Tumorigenesis in Mlh1 and Mlh1/Apc1638N mutant mice. *Cancer Res, 59*(6), 1301-1307.
Edelmann, W., Yang, K., Umar, A., Heyer, J., Lau, K., Fan, K., . . . Kucherlapati, R. (1997). Mutation in the mismatch repair gene Msh6 causes cancer susceptibility. *Cell, 91*(4), 467-477. doi:10.1016/s0092-8674(00)80433-x
Egashira, A., Yamauchi, K., Yoshiyama, K., Kawate, H., Katsuki, M., Sekiguchi, M., . . . Tsuzuki, T. (2002). Mutational specificity of mice defective in the MTH1 and/or the MSH2 genes. *DNA Repair (Amst), 1*(11), 881-893. doi:10.1016/s1568-7864(02)00113-1
Ellegren, H. (2000). Heterogeneous mutation processes in human microsatellite DNA sequences. *Nat Genet, 24*(4), 400-402. doi:10.1038/74249
Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nat Rev Genet, 5*(6), 435-445. doi:10.1038/nrg1348
Fan, H., & Chu, J. Y. (2007). A brief review of short tandem repeat mutation. *Genomics Proteomics Bioinformatics, 5*(1), 7-14. doi:10.1016/S1672-0229(07)60009-6
Fayomi, A. P., & Orwig, K. E. (2018). Spermatogonial stem cells and spermatogenesis in mice, monkeys and men. *Stem Cell Res, 29*, 207-214. doi:10.1016/j.scr.2018.04.009

Fennelly, J., Wright, E., & Plumb, M. (1997). Mini- and microsatellite mutations in radiation-induced acute myeloid leukaemia in the CBA/H mouse. *Leukemia, 11*(6), 807-810.
Fraga, M. F., Herranz, M., Espada, J., Ballestar, E., Paz, M. F., Ropero, S., . . . Esteller, M. (2004). A mouse skin multistage carcinogenesis model reflects the aberrant DNA methylation patterns of human tumors. *Cancer Res, 64*(16), 5527-5534. doi:10.1158/0008-5472.CAN-03-4061
Frigola, J., Sabarinathan, R., Mularoni, L., Muinos, F., Gonzalez-Perez, A., & Lopez-Bigas, N. (2017). Reduced mutation rate in exons due to differential mismatch repair. *Nat Genet*. doi:10.1038/ng.3991
Fu, Z., Regan, K., Zhang, L., Muders, M. H., Thibodeau, S. N., French, A., . . . Tindall, D. J. (2009). Deficiencies in Chfr and Mlh1 synergistically enhance tumor susceptibility in mice. *J Clin Invest, 119*(9), 2714-2724. doi:10.1172/JCI37405
Ganai, R. A., & Johansson, E. (2016). DNA Replication-A Matter of Fidelity. *Mol Cell, 62*(5), 745-755. doi:10.1016/j.molcel.2016.05.003
Gazzoli, I., Loda, M., Garber, J., Syngal, S., & Kolodner, R. D. (2002). A hereditary nonpolyposis colorectal carcinoma case associated with hypermethylation of the MLH1 gene in normal tissue and loss of heterozygosity of the unmethylated allele in the resulting microsatellite instability-high tumor. *Cancer Res, 62*(14), 3925-3928.
Goel, A., Nguyen, T. P., Leung, H. C., Nagasaka, T., Rhees, J., Hotchkiss, E., . . . Hitchins, M. P. (2011). De novo constitutional MLH1 epimutations confer early-onset colorectal cancer in two new sporadic Lynch syndrome cases, with derivation of the epimutation on the paternal allele in one. *Int J Cancer, 128*(4), 869-878. doi:10.1002/ijc.25422
Gunes, S., Al-Sadaan, M., & Agarwal, A. (2015). Spermatogenesis, DNA damage and DNA repair mechanisms in male infertility. *Reprod Biomed Online, 31*(3), 309-319. doi:10.1016/j.rbmo.2015.06.010
Hampel, H., Frankel, W. L., Martin, E., Arnold, M., Khanduja, K., Kuebler, P., . . . de la Chapelle, A. (2005). Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). *N Engl J Med, 352*(18), 1851-1860. doi:10.1056/NEJMoa043146
Hemminki, A., Peltomaki, P., Mecklin, J. P., Jarvinen, H., Salovaara, R., Nystrom-Lahti, M., . . . Aaltonen, L. A. (1994). Loss of the wild type MLH1 gene is a feature of hereditary nonpolyposis colorectal cancer. *Nat Genet, 8*(4), 405-410. doi:10.1038/ng1294-405
Herman, J. G., Graff, J. R., Myohanen, S., Nelkin, B. D., & Baylin, S. B. (1996). Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A, 93*(18), 9821-9826. doi:10.1073/pnas.93.18.9821
Hitchins, M. P. (2013). The role of epigenetics in Lynch syndrome. *Fam Cancer, 12*(2), 189-205. doi:10.1007/s10689-013-9613-3
Hitchins, M. P., Wong, J. J., Suthers, G., Suter, C. M., Martin, D. I., Hawkins, N. J., & Ward, R. L. (2007). Inheritance of a cancer-associated MLH1 germ-line epimutation. *N Engl J Med, 356*(7), 697-705. doi:10.1056/NEJMoa064522
Holliday, R., & Pugh, J. E. (1975). DNA modification mechanisms and gene activity during development. *Science, 187*(4173), 226-232.
Jones, P. A., & Baylin, S. B. (2007). The epigenomics of cancer. *Cell, 128*(4), 683-692. doi:10.1016/j.cell.2007.01.029
Kabbarah, O., Mallon, M. A., Pfeifer, J. D., Edelmann, W., Kucherlapati, R., & Goodfellow, P. J. (2003). A panel of repeat markers for detection of microsatellite instability in murine tumors. *Mol Carcinog, 38*(4), 155-159. doi:10.1002/mc.10157
Kanaya, T., Kyo, S., Maida, Y., Yatabe, N., Tanaka, M., Nakamura, M., & Inoue, M. (2003). Frequent hypermethylation of MLH1 promoter in normal endometrium of patients with endometrial cancers. *Oncogene, 22*(15), 2352-2360. doi:10.1038/sj.onc.1206365
Kansikas, M., Kasela, M., Kantelinen, J., & Nystrom, M. (2014). Assessing how reduced expression levels of the mismatch repair genes MLH1, MSH2, and MSH6 affect repair efficiency. *Hum Mutat, 35*(9), 1123-1127. doi:10.1002/humu.22605

Kawate, H., Sakumi, K., Tsuzuki, T., Nakatsuru, Y., Ishikawa, T., Takahashi, S., . . . Sekiguchi, M. (1998). Separation of killing and tumorigenic effects of an alkylating agent in mice defective in two of the DNA repair genes. *Proc Natl Acad Sci U S A, 95*(9), 5116-5120. doi:10.1073/pnas.95.9.5116
Kneitz, B., Cohen, P. E., Avdievich, E., Zhu, L., Kane, M. F., Hou, H., Jr., . . . Edelmann, W. (2000). MutS homolog 4 localization to meiotic chromosomes is required for chromosome pairing during meiosis in male and female mice. *Genes Dev, 14*(9), 1085-1097.
Knudson, A. G., Jr. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A, 68*(4), 820-823. doi:10.1073/pnas.68.4.820
Komissarov, A. S., Gavrilova, E. V., Demin, S. J., Ishov, A. M., & Podgornaya, O. I. (2011). Tandemly repeated DNA families in the mouse genome. *BMC Genomics, 12*, 531. doi:10.1186/1471-2164-12-531
Laird, C. D. (1971). Chromatid structure: relationship between DNA content and nucleotide sequence diversity. *Chromosoma, 32*(4), 378-406. doi:10.1007/BF00285251
Lee, K., Tosti, E., & Edelmann, W. (2016). Mouse models of DNA mismatch repair in cancer research. *DNA Repair (Amst), 38*, 140-146. doi:10.1016/j.dnarep.2015.11.015
Li, H., & Jasper, H. (2016). Gastrointestinal stem cells in health and disease: from flies to humans. *Dis Model Mech, 9*(5), 487-499. doi:10.1242/dmm.024232
Lin, D. P., Wang, Y., Scherer, S. J., Clark, A. B., Yang, K., Avdievich, E., . . . Edelmann, W. (2004). An Msh2 point mutation uncouples DNA mismatch repair and apoptosis. *Cancer Res, 64*(2), 517-522. doi:10.1158/0008-5472.can-03-2957
Lipkin, S. M., Moens, P. B., Wang, V., Lenzi, M., Shanmugarajah, D., Gilgeous, A., . . . Cohen, P. E. (2002). Meiotic arrest and aneuploidy in MLH3-deficient mice. *Nat Genet, 31*(4), 385-390. doi:10.1038/ng931
Long, M. D., Smiraglia, D. J., & Campbell, M. J. (2017). The Genomic Impact of DNA CpG Methylation on Gene Expression; Relationships in Prostate Cancer. *Biomolecules, 7*(1). doi:10.3390/biom7010015
Lu, J. Y., Shao, W., Chang, L., Yin, Y., Li, T., Zhang, H., . . . Shen, X. (2020). Genomic Repeats Categorize Genes with Distinct Functions for Orchestrated Regulation. *Cell Rep, 30*(10), 3296-3311 e3295. doi:10.1016/j.celrep.2020.02.048
Lynch, H. T., & de la Chapelle, A. (2003). Hereditary colorectal cancer. *N Engl J Med, 348*(10), 919-932. doi:10.1056/NEJMra012242
Markowitz, S., Wang, J., Myeroff, L., Parsons, R., Sun, L., Lutterbaugh, J., . . . et al. (1995). Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science, 268*(5215), 1336-1338.
McCulloch, S. D., & Kunkel, T. A. (2008). The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res, 18*(1), 148-161. doi:10.1038/cr.2008.4
Mead, L. J., Jenkins, M. A., Young, J., Royce, S. G., Smith, L., St John, D. J., . . . Southey, M. C. (2007). Microsatellite instability markers for identifying early-onset colorectal cancers caused by germ-line mutations in DNA mismatch repair genes. *Clin Cancer Res, 13*(10), 2865-2869. doi:10.1158/1078-0432.CCR-06-2174
Miyakura, Y., Sugano, K., Akasu, T., Yoshida, T., Maekawa, M., Saitoh, S., . . . Nagai, H. (2004). Extensive but hemiallelic methylation of the hMLH1 promoter region in early-onset sporadic colon cancers with microsatellite instability. *Clin Gastroenterol Hepatol, 2*(2), 147-156. doi:10.1016/s1542-3565(03)00314-8
Moolenbeek, C., & Ruitenberg, E. J. (1981). The "Swiss roll": a simple technique for histological studies of the rodent intestine. *Lab Anim, 15*(1), 57-59.
Morak, M., Schackert, H. K., Rahner, N., Betz, B., Ebert, M., Walldorf, C., . . . Holinski-Feder, E. (2008). Further evidence for heritability of an epimutation in one of 12 cases with MLH1 promoter methylation in blood cells clinically displaying HNPCC. *Eur J Hum Genet, 16*(7), 804-811. doi:10.1038/ejhg.2008.25
Mukherjee, S., Ridgeway, A. D., & Lamb, D. J. (2010). DNA mismatch repair and infertility. *Curr Opin Urol, 20*(6), 525-532. doi:10.1097/MOU.0b013e32833f1c21
Niessen, R. C., Hofstra, R. M., Westers, H., Ligtenberg, M. J., Kooi, K., Jager, P. O., . . . Sijmons, R. H. (2009). Germline hypermethylation of MLH1 and EPCAM deletions are a frequent cause of Lynch syndrome. *Genes Chromosomes Cancer, 48*(8), 737-744. doi:10.1002/gcc.20678

Ohta, T., & Kimura, M. (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet Res, 22*(2), 201-204. doi:10.1017/s0016672300012994
Parsons, R., Li, G. M., Longley, M., Modrich, P., Liu, B., Berk, T., . . . Vogelstein, B. (1995). Mismatch repair deficiency in phenotypically normal human cells. *Science, 268*(5211), 738-740.
Pedroni, M., Tamassia, M. G., Percesepe, A., Roncucci, L., Benatti, P., Lanza, G., Jr., . . . Ponz de Leon, M. (1999). Microsatellite instability in multiple colorectal tumors. *Int J Cancer, 81*(1), 1-5. doi:10.1002/(sici)1097-0215(19990331)81:1<1::aid-ijc1>3.0.co;2-k
Peltomaki, P. (2014). Epigenetic mechanisms in the pathogenesis of Lynch syndrome. *Clin Genet, 85*(5), 403-412. doi:10.1111/cge.12349
Pray, L. A. (2008). DNA Replication and Causes of Mutation. *Nature Education, 1*(1), 214.
Prolla, T. A., Baker, S. M., Harris, A. C., Tsao, J. L., Yao, X., Bronner, C. E., . . . Liskay, R. M. (1998). Tumour susceptibility and spontaneous mutation in mice deficient in Mlh1, Pms1 and Pms2 DNA mismatch repair. *Nat Genet, 18*(3), 276-279. doi:10.1038/ng0398-276
Reitmair, A. H., Schmits, R., Ewel, A., Bapat, B., Redston, M., Mitri, A., . . . et al. (1995). MSH2 deficient mice are viable and susceptible to lymphoid tumours. *Nat Genet, 11*(1), 64-70. doi:10.1038/ng0995-64
Richard, G. F., Kerrest, A., & Dujon, B. (2008). Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev, 72*(4), 686-727. doi:10.1128/MMBR.00011-08
Rogacheva, M. V., Manhart, C. M., Chen, C., Guarne, A., Surtees, J., & Alani, E. (2014). Mlh1-Mlh3, a meiotic crossover and DNA mismatch repair factor, is a Msh2-Msh3-stimulated endonuclease. *J Biol Chem, 289*(9), 5664-5673. doi:10.1074/jbc.M113.534644
Ryland, G. L., Doyle, M. A., Goode, D., Boyle, S. E., Choong, D. Y., Rowley, S. M., . . . Gorringe, K. L. (2015). Loss of heterozygosity: what is it good for? *BMC Med Genomics, 8*, 45. doi:10.1186/s12920-015-0123-z
Sancho, R., Cremona, C. A., & Behrens, A. (2015). Stem cell and progenitor fate in the mammalian intestine: Notch and lateral inhibition in homeostasis and disease. *EMBO Rep, 16*(5), 571-581. doi:10.15252/embr.201540188
Saxonov, S., Berg, P., & Brutlag, D. L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A, 103*(5), 1412-1417. doi:10.1073/pnas.0510310103
Shah, S. N., Hile, S. E., & Eckert, K. A. (2010). Defective mismatch repair, microsatellite mutation bias, and variability in clinical cancer phenotypes. *Cancer Res, 70*(2), 431-435. doi:10.1158/0008-5472.CAN-09-3049
Shao, C., Deng, L., Chen, Y., Kucherlapati, R., Stambrook, P. J., & Tischfield, J. A. (2004). Mlh1 mediates tissue-specific regulation of mitotic recombination. *Oncogene, 23*(56), 9017-9024. doi:10.1038/sj.onc.1208148
Suter, C. M., Martin, D. I., & Ward, R. L. (2004). Germline epimutation of MLH1 in individuals with multiple cancers. *Nat Genet, 36*(5), 497-501. doi:10.1038/ng1342
Tabori, U., Hansford, J. R., Achatz, M. I., Kratz, C. P., Plon, S. E., Frebourg, T., & Brugieres, L. (2017). Clinical Management and Tumor Surveillance Recommendations of Inherited Mismatch Repair Deficiency in Childhood. *Clin Cancer Res, 23*(11), e32-e37. doi:10.1158/1078-0432.CCR-17-0574
Thiagalingam, S., Laken, S., Willson, J. K., Markowitz, S. D., Kinzler, K. W., Vogelstein, B., & Lengauer, C. (2001). Mechanisms underlying losses of heterozygosity in human colorectal cancers. *Proc Natl Acad Sci U S A, 98*(5), 2698-2702. doi:10.1073/pnas.051625398
Thibodeau, S. N., Bren, G., & Schaid, D. (1993). Microsatellite instability in cancer of the proximal colon. *Science, 260*(5109), 816-819. doi:10.1126/science.8484122
Thomas, D. C., Roberts, J. D., Sabatino, R. D., Myers, T. W., Tan, C. K., Downey, K. M., . . . Kunkel, T. A. (1991). Fidelity of mammalian DNA replication and replicative DNA polymerases. *Biochemistry, 30*(51), 11751-11759. doi:10.1021/bi00115a003
Tokairin, Y., Kakinuma, S., Arai, M., Nishimura, M., Okamoto, M., Ito, E., . . . Shimada, Y. (2006). Accelerated growth of intestinal tumours after radiation exposure in Mlh1-knockout mice: evaluation

of the late effect of radiation on a mouse model of HNPCC. *Int J Exp Pathol, 87*(2), 89-99. doi:10.1111/j.0959-9673.2006.00464.x
Treangen, T. J., & Salzberg, S. L. (2011). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet, 13*(1), 36-46. doi:10.1038/nrg3117
Umar, A., Boland, C. R., Terdiman, J. P., Syngal, S., de la Chapelle, A., Ruschoff, J., . . . Srivastava, S. (2004). Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst, 96*(4), 261-268. doi:10.1093/jnci/djh034
Umar, A., Risinger, J. I., Hawk, E. T., & Barrett, J. C. (2004). Testing guidelines for hereditary non-polyposis colorectal cancer. *Nat Rev Cancer, 4*(2), 153-158. doi:10.1038/nrc1278
Valo, S., Kaur, S., Ristimaki, A., Renkonen-Sinisalo, L., Jarvinen, H., Mecklin, J. P., . . . Peltomaki, P. (2015). DNA hypermethylation appears early and shows increased frequency with dysplasia in Lynch syndrome-associated colorectal adenomas and carcinomas. *Clin Epigenetics, 7*, 71. doi:10.1186/s13148-015-0102-4
Varghese, F., Bukhari, A. B., Malhotra, R., & De, A. (2014). IHC Profiler: an open source plugin for the quantitative evaluation and automated scoring of immunohistochemistry images of human tissue samples. *PLoS One, 9*(5), e96801. doi:10.1371/journal.pone.0096801
Vasen, H. F., Watson, P., Mecklin, J. P., & Lynch, H. T. (1999). New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. *Gastroenterology, 116*(6), 1453-1456. doi:10.1016/s0016-5085(99)70510-x
Watson, P., & Riley, B. (2005). The tumor spectrum in the Lynch syndrome. *Fam Cancer, 4*(3), 245-248. doi:10.1007/s10689-004-7994-z
Wei, K., Kucherlapati, R., & Edelmann, W. (2002). Mouse models for human DNA mismatch-repair gene defects. *Trends Mol Med, 8*(7), 346-353.
Wijnen, J. T., Vasen, H. F., Khan, P. M., Zwinderman, A. H., van der Klift, H., Mulder, A., . . . Fodde, R. (1998). Clinical findings with implications for genetic testing in families with clustering of colorectal cancer. *N Engl J Med, 339*(8), 511-518. doi:10.1056/NEJM199808203390804
Vilkki, S., Launonen, V., Karhu, A., Sistonen, P., Vastrik, I., & Aaltonen, L. A. (2002). Screening for microsatellite instability target genes in colorectal cancers. *J Med Genet, 39*(11), 785-789.
Yang, G., Scherer, S. J., Shell, S. S., Yang, K., Kim, M., Lipkin, M., . . . Edelmann, W. (2004). Dominant effects of an Msh6 missense mutation on DNA repair and cancer susceptibility. *Cancer Cell, 6*(2), 139-150. doi:10.1016/j.ccr.2004.06.024
Yauk, C. L., Dubrova, Y. E., Grant, G. R., & Jeffreys, A. J. (2002). A novel single molecule analysis of spontaneous and radiation-induced mutation at a mouse tandem repeat locus. *Mutat Res, 500*(1-2), 147-156.
Zhang, Y., Monckton, D. G., Siciliano, M. J., Connor, T. H., & Meistrich, M. L. (2002). Detection of radiation and cyclophosphamide-induced mutations in individual mouse sperm at a human expanded trinucleotide repeat locus transgene. *Mutat Res, 516*(1-2), 121-138. doi:10.1016/s1383-5718(02)00035-9